

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН  
НЕКОММЕРЧЕСКОЕ АКЦИОНЕРНОЕ ОБЩЕСТВО  
«АЛМАТИНСКИЙ УНИВЕРСИТЕТ ЭНЕРГЕТИКИ И СВЯЗИ ИМЕНИ ГУМАРБЕКА  
ДАУКЕЕВА»

ISSN 2790-0886

# **В Е С Т Н И К**

## **АЛМАТИНСКОГО УНИВЕРСИТЕТА ЭНЕРГЕТИКИ И СВЯЗИ**

Учрежден в июне 2008 года

Тематическая направленность: энергетика и энергетическое машиностроение, информационные,  
телекоммуникационные и космические технологии

---

**2 (57)**

**2022**

Импакт-фактор - 0.105

Научно-технический журнал  
Выходит 4 раза в год

Алматы

**СВИДЕТЕЛЬСТВО**

о постановке на переучет периодического печатного издания,  
информационного агентства и сетевого издания  
№ KZ14VPY00024997

выдано

Министерством информации и общественного развития  
Республики Казахстан

Подписной индекс – 74108

---

**Бас редакторы – главный редактор**

Стояк В.В.

к. т.н., профессор

Заместитель главного редактора  
Ответственный секретарь

Жауыт Алгазы, доктор PhD  
Шуебаева Д.А., магистр

**Редакция алқасы – Редакционная коллегия**

**Главный редактор – Стояк В.В.**, кандидат технических наук, профессор Алматинского Университета Энергетики и Связи имени Гумарбека Даукеева, Казахстан;

**Заместитель главного редактора – Жауыт А.**, доктор PhD, ассоциированный профессор Алматинского Университета Энергетики и Связи имени Гумарбека Даукеева, Казахстан;

**Сагинтаева С.С.**, ректор Алматинского Университета Энергетики и Связи имени Гумарбека Даукеева, доктор экономических наук, кандидат физико-математических наук, профессор математики, академик МАИН;

**Ревалде Г.**, доктор PhD, член-корреспондент Академии наук, директор Национального Совета науки, Рига, Латвия;

**Илиев И.К.**, доктор технических наук, Русенский университет, Болгария;

**Белоев К.**, доктор технических наук, профессор Русенского университета, Болгария;

**Обозов А.Д.**, доктор технических наук, НАН Кыргызской Республики, заведующий Лабораторией «Возобновляемые источники энергии», Кыргызская Республика;

**Кузнецов А.А.**, доктор технических наук, профессор Омского государственного технического университета, ОмГУПС, Российская Федерация, г. Омск;

**Алипбаев К.А.**, PhD, доцент Алматинского Университета Энергетики и Связи имени Гумарбека Даукеева, Казахстан;

**Зверева Э.Р.**, доктор технических наук, профессор Казанского государственного энергетического университета, Российская Федерация, г. Казань;

**Лахно В.А.**, доктор технических наук, профессор Национального университета биоресурсов и природопользования Украины, кафедра компьютерных систем, сетей и кибербезопасности, Украина, Киев;

**Омаров Ч.Т.**, кандидат физико-математических наук, директор Астрофизического института имени В.Г. Фесенкова, Казахстан;

**Коньшин С.В.**, кандидат технических наук, профессор Алматинского Университета Энергетики и Связи имени Гумарбека Даукеева, Казахстан;

**Тынымбаев С.Т.**, кандидат технических наук, профессор Алматинского Университета Энергетики и Связи имени Гумарбека Даукеева, Казахстан.



## ИНФОРМАЦИОННЫЕ, ТЕЛЕКОММУНИКАЦИОННЫЕ И КОСМИЧЕСКИЕ ТЕХНОЛОГИИ

ОӘҚ 004.8

[https://doi.org/10.51775/2790-0886\\_2022\\_57\\_2\\_70](https://doi.org/10.51775/2790-0886_2022_57_2_70)

### МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНА ОТЫРЫП ДЫБЫСТЫ ЖІКТЕУ

Ж.М. Досбаев<sup>1,2\*</sup>, Б.С. Омаров<sup>2</sup>, Б.М. Садықова<sup>2</sup>, Н.Қ. Смайлов<sup>1</sup>

<sup>1</sup>Сәтбаев университеті, Алматы, Қазақстан

<sup>2</sup>Әл-Фараби атындағы Қазақ Ұлттық университеті, Алматы, Қазақстан

E-mail: [Zh.dosbayev@satbayev.university](mailto:Zh.dosbayev@satbayev.university), [d\\_jandos\\_93@mail.ru](mailto:d_jandos_93@mail.ru), [batyahan@gmail.com](mailto:batyahan@gmail.com),  
[mukhtarkyzy.bibi@gmail.com](mailto:mukhtarkyzy.bibi@gmail.com), [nur\\_aly.kz@mail.ru](mailto:nur_aly.kz@mail.ru)

**Аңдатпа.** Зерттеу жұмысының мақсаты аудиосенсорлардан алынған мәліметтерге аудио жіктеу жүргізуде қолданылатын машиналық оқыту әдістерін пайдалана отырып, оқытылатын модельдің жіктелу нақтылығын арттыру. Аудио деректерді жіктеу кезеңдеріне жеке тоқталып, өзге авторлардың шешімдеріне талдау жүргізілді. Сонымен қатар, белгілерді алу әдістері, ерекшеліктері мен артықшылықтары, теңдеулері және қолданылуы анықталды. Алынған белгілер бойынша машиналық оқыту алгоритмдерін жүргізу, архитектуралары, сондай-ақ оларды дамыту және модельдерді оқыту, тестілеу нәтижелері сипатталды. Сонымен қатар қолданылған аудио data set-тердің ерекшеліктері, жалпы сипаттамалары, класстары қарастырылды. Алдын-ала аудио мәліметтерді өңдеу, сүзгілеу ерекшеліктері, аудио мәліметтердің құрылымдық бірліктері аталды. Алдыңғы зерттеу жұмыстарында дәстүрлі оқыту алгоритмдері бойынша модельге оқыту жүргізілген болса, ал бұл жұмыстың өзектілігі терең оқыту әдістерін қолдану арқылы, яғни нейрондық желілер негізінде бинарлы жіктеу үшін қарапайым модель және мультижіктеу тапсырмасы үшін VGG моделі оқытылды. Зерттеу нәтижесінде нейрондық желіні қолдану арқылы бинарлы жіктелу нақтылығы 98,8% және мультижіктеу тапсырмасы бойынша 92,6% нақтылық алынды. Ал VGG моделінің нақтылығы 74%-дан 85%-ға дейін артты. Машиналық оқыту әдістері негізінде жоғары нақтылыққа ие модельдерді қолдану қоғамдық орындарда және ғимарат ішіндегі қауіпсіздікті қамтамасыз етуге мүмкіндік береді. Жұмыстың жаңалығы ретінде қолданылатын Dataset пен алгоритмдердің тиімділігін арттыру үшін қолданылатын аудиофайлдарды түрлендіру әдістерінің өзгешелігін, алынған белгілер мен алгоритмдердің әр алуандығын, сонымен қатар неғұрлым жоғары нақтылыққа қол жеткізуді атап өтуге болады. Жұмысты әрі қарай дамыту бойынша бағыттар ұсынылады.

**Түйін сөздер:** аудиоклассификация, машиналық оқыту алгоритмдері, аудиоаналитика, нейрондық желі, белгілерді алу.

#### КІРІСПЕ

Сандық ақпараттар дәуірінде, аудио мәліметтер заманауи компьютерлік қосымшалардың маңызды ажырамас бөлігі болып табылады. Қауіпсіздікті қамтамасыз етуге көбінесе бақылау бейнекамералары қолданылатынына қарамастан, аудиоаналитика көптеген артықшылықтарға ие: қолданылатын құрылғылардың және қызмет көрсету бағасының төмендігі; ақпарат сол мезетте таратылатындықтан аудио мәліметтердің көлемінің аз болуы, тарату арналарының өткізу қабілетіне қойылатын талаптарды жеңілдетеді. Әсіресе қала қауіпсіздігін қамтамасыз ету үшін, күдікті дыбыс анықталған аумаққа назар аударуда және орын алып жатқан оқиғаларды анықтауда, аудиоаналитиканы пайдалану тиімділігі тақырыптың өзектілігін айқындайды.

Аудио дыбыстардың жіктелуі немесе категорияға бөлінуі негізгі екі қадамнан тұрады. Бірінші қадам, дыбыстың белгілерін (feature extraction) алу әдістерін қолдана отырып, аудио дыбыстың шағын параметрлері алынады, ал екінші қадам, алынған параметрлер бойынша қарапайым Евклидтік аралық әдістен бастап, күрделі статистикалық әдістерге дейін қолданып, жіктеу мен категорияға бөлу жүргізіледі. Аудио жіктеу және категорияға бөлу жүйесінің тиімділігі аудио белгілердің (feature)

дұрыс алынуына және әрбір белгілер жиынтығын тиісті классқа жіктеу нақтылығына байланысты болады.

### ӘДЕБИЕТТЕРГЕ ШОЛУ

Датчиктен алынған аудио мәліметтер машиналық оқыту процесіне жіберілуден бұрын, алдын-ала өңдеуді талап етеді. Бұл өңдеу бөгеуіл мен шуылды басу үшін сүзгілеуден немесе принциптік компоненттік анализ (PCA) немесе мәлімет өлшемін азайту үшін тәуелсіз компоненттік анализ (ICA) жүргізуден тұрады [1], [2]. Тиімді модельдерді генерациялауда және тиімді оқытуды қамтамасыз етуде, көптеген оқыту алгоритмдері үшін белгілерді алу және таңдау өте күрделі әрі маңызды тапсырма болып табылады.

Rong ұсынған зерттеу жұмысында иерархиялық құрылым төрт деңгейден тұратын audio data-ны қолдана отырып, тірек векторлары әдісін (SVM) Гаусс ядросы (Gaussian kernel) арқылы жақсарту негізінде жіктеу жүргізеді [3]. Сонымен қатар, жұмыста, әр түрлі уақыт бөлігінде бар аудионың құрылымдық бірліктерінің анықтамалары ұсынылады:

- Аудио кадр (audio frame): аудио – кездейсоқ әрі тұрақсыз процесс және сипаттамалары уақыттың өтуімен өзгеріп отырады, алайда, бұл сипаттамалар баяу өзгереді. Сондықтан дыбыс сигналын қысқа өңдеу периодтары қатарларына бөлуге болады. Қысқа сегменттердің ұзақтығы 20-30 мс және аудио кадр дыбыс өңдеудегі ең кіші бірлік болып табылады;
- Аудио клип: аудио кадрлардың уақыт бірлігі өте кіші және семантикалық құрамын алу өте қиын болғандықтан, уақыт бірлігі біршама үлкен болатын жаңа аудио бірлігі (audio clip) арқылы анықтау қажет болады. Аудио клиптер, ұзақтығы белгіленген, бірнеше кадрлардан тұрады. Аудио сегменттердің сипаттамалары аудио кадрлар негізінде есептеледі;
- Аудио бөлшек (audio shot): Бұл ұғым бейнематериалдан келіп шығады. Аудио сегмент өте қысқа болғандықтан, семантикалық құрам анализінде қолдануға ыңғайлы емес. Аудионың құрылымдық бірлігінің құрамында бұл класс аудио бөлшек деп анықталады. Аудио бөлшек, бір классқа жататын, бірнеше аудио сегменттерден құралады.
- Жоғары деңгейлі семантикалық аудио бірлік: бірнеше аудио бөлшектердің түрлі комбинацияларынан тұратын, семантикалық құрылымы толық, аудио құрылымдық бірлік.

Аудио мәліметтерді (audio data) жіктеу аудио мәліметтің белгілерін алу арқылы орындалады. Аталған жұмыста үш түрлі аудио белгілер (audio features) алынады:

- Short time energy (STE)  $N$  сигнал үлгілерінің квадраттарының қосындысы ретінде анықталатын  $x[n]$  дискреттік сигнал:

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

- Zero crossing rate (ZCR) функция белгісінің өзгеру саны туралы ақпарат береді (функцияның ОХ осін қиып өтуі). Егер қиып өту саны үлкен болса, демек сигнал жоғары жиілікті ақпаратты құрайды немесе керісінше. Осылайша ZCR сигналдың жиіліктік құрамын туралы ақпаратты және шуылды өлшемін ұсынады:

$$F(i) = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1} \sum_{k=1}^{n_+} (x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1} \sum_{k=1}^{n_-} (x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \quad (2)$$

- Mel-frequency cepstral (MFCC) коэффициенті дискреттік косинустық түрлендіру негізінде анықталады. Бұл анализдің түрі дыбыс классификациясы тапсырмаларында, әсіресе, адам сөзі дыбысын тану жүйелерінде кеңінен қолданыс тапқан [4]. Бұл коэффициентінің негізгі шкаласы адамның перцептивті жүйесі болып табылады. MFCC алу акустикалық сигналдың келесі сатыларға сәйкес өңдеуін жүргізеді: сигналды кадрлерге бөлу және Хемминг терезесі функциясын қолдану; әр

кадр үшін амплитудалық спектр алу; спектрлердің логарифмін алу; Mel шкалаға түрлендіру; дискреттік косинустық түрлендіруді орындау:

$$c_n = \sqrt{\frac{2}{K}} \cdot \sum_{k=1}^K (\log S_k) \cdot \cos\left(\frac{1}{K} \cdot (n(k-0.5)\pi)\right) \quad (3)$$

Мұндағы:  $K$  – жолақтық сүзгілер саны.

SVM жіктегіші ықшамдау тапсырмасын төмендегідей шешуге арналған:

$$\min_{\omega, b, \xi} \left( \frac{1}{2} \omega^T \omega + C \sum_i \xi_i \right) \quad (4)$$

Аудио жіктеуді жүргізу нәтижесінде автор келесі шешімді алған:

$$\text{sgn}(\omega^T \phi(x) + b) = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + b\right) \quad (5)$$

SVM алгоритмінің өнімділігін арттыру үшін Гаусс ядросын жіктегішке қолданады:

$$K(x_i, x) = \exp\left(-\frac{1}{\sigma^2} \|x - x_i\|^2\right) \quad (6)$$

Нәтижесінде GS (general sounds) және AS (audio scenes) мәліметтер базалары бойынша орташа жіктеу нақтылығы (average classification accuracy) тиісінше 0.876 және 0.863 анықталды. Авторлар, алдағы уақытта аталған алгоритмді, басқа dataset-ге қолдануды ұсынған.

Josef, Eva және т.б. авторлар SVM-BT (support vector machine – binary tree) архитектурасын пайдалана отырып, BN (broadcast news) audio data-сы негізгі media келесі 5 санат бойынша жіктеу жүргізеді [5]: pure speech - тыныш ортада немесе студиядағы репортаж, яғни еш шусыз фондағы адам дыбысы; speech with environment sound – шулы ортадағы дыбыстар, интервью және телефондағы сөйлесулер; speech with music background - жаңалықтардың басындағы және соңындағы әуендер, адамның сөйлеуімен қоса жарнамалық роликтер; pure music – әуендер, жарнамалық роликтердегі музыкалар; environment sound – аудиоағындағы акустикалық оқиғалардың барлық түрлері, мысалы, көліктің құстардың, судың, желдің, топтың дауыстары және т.б.; silence - динамиктер арасындағы және роликтердегі тыныш интервалдар.

Database ретінде Slovak TV аудиоағындарының жазбаларынан тұратын KEMT-BN1 database-і қолданылады [6]. Аудиофайлдар 48 кГц 192 кБит стерео MPEG Layer 2 аудиоформатында алынған. MPEG файлдар PCM 16 кГц 16 Бит моно форматты қолдану арқылы WAV файлга конвертацияланып, белгілерді (feature) алуға қолданылады.

Аталған Database жалпы ұзақтығы 65 сағат болатын 188 жазбадан тұрады. Тек дыбыстан тұратын кадрлардың (frame) ұзақтығы 55 сағат. Ал еш шуылсыз таза дыбыстардан тұратын кадрларды қарастыратын болсақ, онда жалпы ұзақтық 21 сағат 47 минутты құрайды.

Сегменттеу аудиосигналды алдын-ала өңдеу немесе алдын-ала сегменттеу тапсырмаларымен байланысты. Анығырақ айтқанда, ұзындығы тең сегменттерді анықтау. Яғни, бұл сегменттердің ауытқуы, қысқа уақыттағы энергияның (STE) орташа квадраттық мәні (RMS), нөлдік қию жылдамдығы (ZCR) сияқты уақыттық облыстағы статистикалық параметрлері есептеледі. Қарапайым жағдайда бір сегменттің ұзақтығы 1 секундқа тең. Алдын-ала сегменттеуден соң спектралдық бұрмалануды жою үшін Хемминг терезесін қолдану арқылы overlap кадрларына бөлінеді.

Авторлар BN сегменттеудің және жіктеу жүргізудің негізгі әрі маңызды тапсырмасы, аудиоконтенттерді объективті ұсына алатын, белгілерді (feature) алу екендігін атай кетіп [5], жиі қолданылатын белгілерді келесі төрт классқа жіктейді [7]:

- Уақыттық облыстағы белгілер – қысқа уақыттағы энергиясы (STE), тыныштық кадрларының қатынасы (silence frame ratio – SFR), кадрларының қатынасы (noise frame ratio - NFR), диапазондағы энергияның таралуы (subband energy distribution - SED), энергиясы төмен кадрлардың пайызы

(percentage of low-energy frames) және модуляция энергиясы (modulation energy), нөлдік қию жылдамдығы (zero-crossing rate - ZCR);

- Уақыттық облыстағы/спектрлік белгілер – бұл белгілердің көбісі MPEG-7 стандартында анықталған [8], мысалы, spectral roll-of (ROF), дыбыстық спектр центроиді (audio spectrum centroid - ASC), спектралдық ағын (spectral flux - SF), экцесстің спектралдық коэффициенті (spectral kurtosis - SK), дыбыстық спектрдің таралуы (audio spectrum spread - ASS) және жазықтығы (flatness - ASF) және т.б.;

- Кепстрлік белгілер - Mel Frequency cepstral coefficients (MFCCs), сызықтық болжамдық кепстралдық коэффициент (linear predictive cepstral coefficients - LPCC), сызықтық спектралдық жұптар (linear spectral pairs - LSP) және перцептуалдық сызықтық болжамдық сараптама (perceptual linear predictive analysis - PLP);

- Беріліс (pitch) белгілері – кадам және дельта кадам (delta pitch - DP), жоғарғы нүктенің спектралдық ұзақтығы (spectral peak duration - SPD), кадамның орташа тығыздығы (average pitch-density - APD) және қуаттың тоналдық тығыздығының қатынасы (relative tonal power density - RTPD).

Бұл жұмыста Rong және өзге авторлар ұсынатын жұмыспен салыстырғанда STE, ZCR және MFCC тағы бірнеше аудио белгілерді қарастырады:

- ASS (audio spectrum spread) – спектралдық форма өлшемі және логарифмдік-жиіліктік спектрдің орта моменті болып табылады [9]. Сонымен қатар, ASS, спектрдің өз центроидінің айналасында қалай таралатындығы туралы ақпарат береді. ASS төмен мәні спектрдің центроид аймағына шоғырлануын білдіреді, ал жоғарғы мәні қуаттың неғұрлым жоғары жиілік диапазонында таралуын көрсетеді;

- ASF (Audio Spectrum Flatness)– қуат спектрінің жазықтығын қасиетін көрсетеді [9]. Бұл белгіні сигнал спектрінің жазық спектрден ауытқуы деп қабылдауға болады. Барлық спектр үшін бір жазықтық мәнін есептеудің орнына жиіліктік жолақтарға бөлу орындалады, тиісінше, нәтижесінде кадр үшін жазықтық мәдерінің бір векторы алынады. ASF жоғарғы коэффициенті шуылды көрсетсе, ал төменгі мәні спектр құрылымының үйлесімді екендігін нұсқайды;

- ROF point (spectral roll-of) [9] – амплитуда спектрі таралуының 85% -нан төмен мәніне негізделген жиілік ретінде анықталады. Сонымен қатар, спектралдық форманың өлшемі және жоғары жиіліктер үшін неғұрлым жоғары мәндерді береді. ROF адам сөзінің дауысын үнсіз музыкадан ажырату үшін қолданылады.

Мәліметтер жиынтығындағы барлық класстады ажырату үшін түрлі белгілер біркелкі емес. Белгілердің әмбебап жиынтығы класстарды ажыратуға қиындық туғызады. Сондықтан, әр класс үшін жіктеудің жоғарғы тиімділігіне қол жеткізу мақсатында жүргізілетін, тиімді белгілерді таңдауға арналған әдістер мен алгоритмдер орындалады. Жоғарыда аталған Josef, Eva және өзге авторлар F-score әдісін қолданады.

F-score, екі оқыту мәліметтері жиынтығы өзгешеліктерін өлшейтін қарапайым әдіс.  $x_k$ ,  $k=1, \dots, m$ , оқыту векторларын ескерсек, егер оң және теріс үлгілер саны  $n_+$  және  $n_-$  тең болса, сәйкесінше,  $i$ -ші белгінің F-score мәні:

$$F(i) \equiv \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1} \sum_{k=1}^{n_+} (x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1} \sum_{k=1}^{n_-} (x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \quad (7)$$

мұндағы  $\bar{x}_i$ ,  $\bar{x}_i^{(+)}$ ,  $\bar{x}_i^{(-)}$  оң және теріс үлгілерді қоса есептегендегі  $i$ -ші белгінің мәні. F-score жоғары мәні, белгінің тиімділік өлшемдері тұрғысынан алғанда, дискриминативтілігін білдіреді. Сондықтан да авторлар оңтайлы белгілер жиынтығын алу барысында, F-score мәні жоғары белгілер бірінші таңдалып, SVM негізінде оқыту/болжау жүргізу үшін қолданады [10].

SVM бинарлық жіктеу тапсырмаларын шешуге арналғандықтан көпклассты жіктеу жүргізуде қиындық тудырады. Бірнеше класстарды дискриминациялау бірнеше екілік жіктегіштерді біріктіру арқылы жүзеге асырылады. Осы себептен авторлар, SVM әдісінің тиімділігін арттыру үшін бұтақ тәріздес сұлбаны жасауға тырысады.

Тиісінше, SVM неғұрлым дискриминативті таралу облыстарын, мысалы, класс шекарасын немесе маржаны модельдеуі керек. Дискриминантты функция барынша жоғары немесе жұмсақ



шекарасы бар, сызықты гипержазықтықпен бөлу арқылы модельденді. Егер әр класстағы оқытушы мәліметтері кеңістікте сызықты бөлінуге келмесе, SVM алдымен кіріс кеңістікті сызықты емес, көп өлшемді белгілер кеңістігіне түрлендіреді. Бұл түрлендірулер полиномиалды, көпқабатты перцептрондағыдай сигмоидты, радиалды-симметриялық функциясы бар RBF-бейнелеу, яғни Гаусстық және сплайндық функциялар арқылы мүмкін болады. Алынатын гипержазықтық, белгілер жазықтығында, максималды маржиналды дискриминантты функция ретінде оңтайлы түрге ие болады. Авторлар, SVM-BT және SVM-OAO (one-against-one multi-class scheme) архитектураларын салыстыра отырып, 1-кестедегі мәндерге қол жеткізеді, сонымен қатар, алдағы уақытта SVM-BT архитектурасын дамыта отырып, сөйлеу сегменттеуін жүргізудің және белгілерді алудың тиімді алгоритмін жасау қажеттігін көрсетеді[5].

1-кесте – SVM-BT архитектурасының жалпы болжау нақтылығы

Нақтылығы [%]	Таза адам сөйлеуі	Сөйлеу және музыка	Сөйлеу және қоршаған орта дыбысы	Музыка	Қораған ота дыбысы	Орташа мәні
SVM-OAO	47.38	59.95	70.77	49.62	33.58	52.26
SVM-BT	83.52	62.78	84.80	93.4	85.79	82.06

Rajkumar Palaniappan және K. Sundaraj ұсынған еңбекте, авторлар, Rale database-ін қолдана отырып, SVM әдісі негізінде, тыныс жолдарының патологиялық жағдайларын анықтауды жүзеге асырады [11]. Rale database 70-тен аса контактілік акселерометр (EMT25C, Siemens) арқылы жазылып алынған, кеуде клеткасының дыбыстарынан тұрады. Жазбалар үш топқа жіктелген: қалыпты, тыныс алу жолдарының обструкциясы және паренхиматозалық патология. Авторлар, жас ерекшеліктері факторы есебінен тек 68 жазбаны қарастырған: 17 жазба - қалыпты; 26 жазба - тыныс алу жолдарының обструкциясы; 25 жазба - паренхиматозалық патология. Тыныс жолдарының дыбысы жүрек дыбысы және т.б. артефактілерге байланысты шуылдарға ұшырағандықтан, алдын-ала өңдеу сатысында тұрақты тоқтың ығысуын жою үшін 7,5 Гц жоғарғы жиілікті 1 реттік Butterworth фильтрі арқылы және спектрлердің қабаттасуын жою үшін 2,5 кГц төменгі жиілікті 8 реттік Butterworth фильтрі арқылы сүзгілеу жүргізілген. Дискретизация жиілігі 10 кГц құрайды.

Mel Frequency cepstral coefficients (MFCCs) коэффициенті, mel scale filter bank сүзгісі арқылы өңделген, FFT (fast fourier transform) коэффициенттері арқылы анықталған. Алынған белгілер SVM жіктегішінің көмегімен анықталды. Жіктегіш тиімділігі 3-кестеде көрсетілгендей матрицалық әдіс арқылы сарапталады. Өнімділікті арттыру үшін, SVM жіктегішіне 60% оқыту және 40% тестілеу орындалды. Барлығы 25 тәжірибе жасалды. Зерттеу нәтижесі 2-кестеде көрсетілген.

2-кесте - MFCC және SVM жіктегіші қызметін қолдану арқылы, респиратор дыбыстарының жіктелу нақтылығы

Жіктелудің ең жоғарғы мәні	Жіктелудің ең төменгі мәні	Жіктелудің Орта мәні
91.35%	88.65%	90.77%

3-кесте - MFCC және SVM жіктегішін қолдану арқылы респиратор дыбысы жіктелу нәтижелеріне матрицалық сараптама

Ақталды	Болжанған				
		Қалыпты	Обструкция	Паренхима	Нақтылығы
	Қалыпты	16	0	1	94.11%
	Обструкция	2	24	0	92.31%
	Паренхима	2	1	22	88.00%

Ұсынылған әдіс бойынша жіктелудің орташа нақтылығы 90,77% құрады. Жіктегіш жұмысының бағасы: 94,11% - қалыпты; 92,31% - тыныс алу жолдарының обструкциясы; 88,00% - паренхима патологиясы. Авторлар алдағы уақытта Linear Predictive cepstral coefficient (LPCC) белгілері негізінде, күрделі машиналық алгоритмдерді жүргізуді ұсынады.

Сонымен, әдебиеттерге шолу бойынша төмендегідей қорытынды жасауға болады: Rong ұсынған зерттеу жұмысында иерархиялық құрылым төрт деңгейден тұратын audio data-ны қолдана отырып, тірек векторлары әдісін (SVM) Гаусс ядросы (Gaussian kernel) арқылы жақсарту негізінде жіктеу жүргізіп, тиісінше 0.876 және 0.863 жіктеу нақтылығына қол жеткізеді. Авторлар, алдағы уақытта аталған алгоритмді, басқа dataset-ге қолдануды ұсынған. Josef, Eva және т.б. авторлар SVM-BT (support vector machine – binary tree) архитектурасын пайдалана отырып, BN (broadcast news) audio data-сы негізінде жіктеу жүргізіп, 0,820 жіктеу нақтылығын алды. Сонымен қатар, алдағы уақытта SVM-BT архитектурасын дамыта отырып, сөйлеу сегментациясын жүргізудің және белгілерді алудың тиімділігі алгоритмін жасау қажеттігін нұсқайды. Rajkumar Palaniappan және K. Sundaraj ұсынған еңбекте авторлар Rale database қолдана отырып, SVM әдісі негізінде, тыныс жолдарының патологиялық жағдайларын анықтауды жүзеге асырып, 0,907 жіктеу нақтылығын алып отыр. Авторлар алдағы уақытта Linear Predictive cepstral coefficient (LPCC) белгілері негізінде, күрделі машиналық алгоритмдерді жүргізуді ұсынады.

### ҚОЙЫЛҒАН ТАПСЫРМАЛАР

Алдыңғы авторлар ұсынған нәтижелерге шолу жасай отырып, жұмысты орындау үшін төмендегідей тапсырамалар қойылды:

1. Dataset жасау.
2. Аудио файлдардың белгілерін алу.
3. Жіктеу әдістерін салыстыру және қолдану.
4. Үлкен көлемдегі dataset негізінде оқытылған модельді алу.

Жұмыстағы ең бірінші қадам өте маңызды, себебі алгоритмдер жұмысының нәтижесінің нақтылығы, мәліметтер жиынтығын дұрыс таңдауға тікелей тәуелді. Зерттеудің бастапқы деректері ретінде WAV форматына түрлендірілген бірарналы аудиофайлдар жиынтығын қарастырамыз.

Әрбір аудиофайлдың сипаттамаларына байланысты, тиісінше, жазбадағы дыбыстық оқиғалар класы жайында ақпарат сақталатын, вектор тұрғызылады және олар жіктегіштер мен нейрондық желілерде кіріс вектор ретінде қолданылады.

Жіктеу тапсырмасын күрделендіре отырып, бинарлы жіктеуден мультижіктеуге өтеміз. Мысалы, мәліметтер жиынтығын адам дауысына байланысты және басқа дыбыстар деп екі классқа бөліп аламыз. Әрі қарай модельге жаңа класстар қосылады және әр аудиофайл бірнеше оқиға белгілерінен тұратындықтан, тапсырма айтарлықтай күрделенеді.

Оқыту нәтижесінде алынатын соңғы модель, келіп түскен аудиосигналға, нақты уақыт мезетінде жіктеу жүргізуі тиіс. Бағдарламаның нәтижесі – алынған модель негізінде болжанған векторлар класы болып табылады. Неғұрлым тиімді әдісті таңдау үшін алгоритмнің нақтылығы мен орындау уақыты ескерілді.

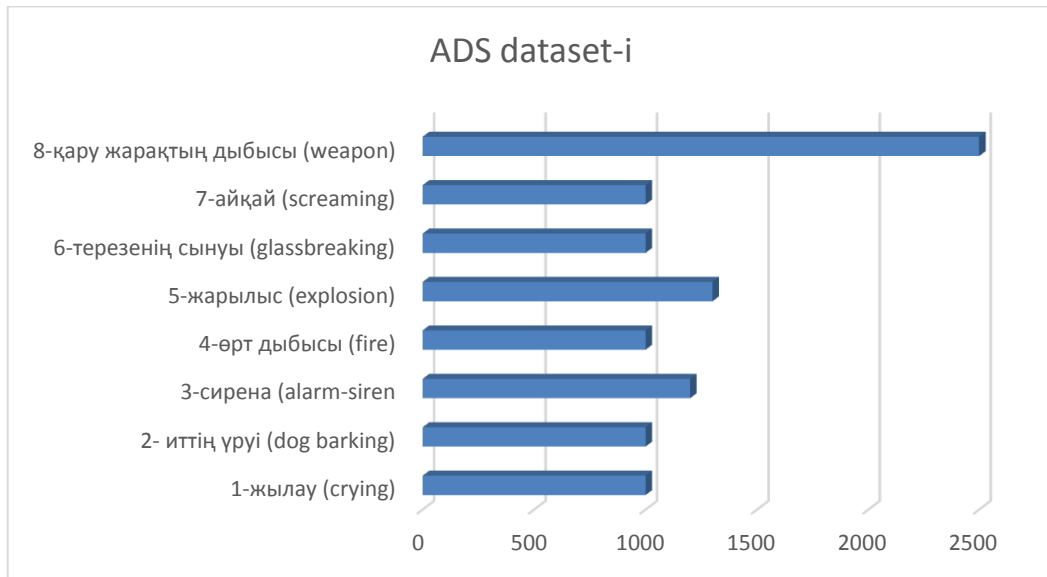
### DATASET

Қоғамдық орындарда қауіпсіздікті қамтамасыз етуде алынатын модельді оқыту үшін 10 мың аудиофайлдан тұратын ADS dataset-і жинақталды. Аудио мәліметтер келесі сегіз класстан тұрады [12]:

- жылау (crying);
- иттің үйі (dog barking);
- сирена (alarm-siren);
- өрт дыбысы (fire);
- жарылыс (explosion);
- терезенің сынуы (glassbreaking);
- айқай (screaming);
- қару жарактың дыбысы (weapon).

Әр класс кем дегенде 1000 аудиофайлдан құралған (1-сурет).





1 - сурет – ADS dataset-індегі класстар бойынша аудиофайлдардың саны

### БЕЛГІЛЕРДІ АЛУ

Әдебиеттерге шолу бөлімінде дыбыстың әр түрлі сипаттамалары қолданылатындығын айтқан болатынбыз. Олар уақытқа, жиілікке және қарқынына (темп) байланысты болуы мүмкін [13], [14]. Жұмыста келесідей аудиосигналдың сипаттамалары алынды.

Энергия – кадр ұзындығы бойынша нормалданған, сигнал мәндері квадратының қосындысы. Мұндағы  $x_i(n)$ ,  $n = 1, \dots, W_L$  деп,  $i$  -ші кадрдағы аудио үлгілердің тізбегі, ал  $W_L$ -кадрдың ұзындығы. Сигнал энергиясының мәні төмендегі формула бойынша анықталады:

$$E(i) = \frac{1}{W_L} \sum_{n=1}^{W_L} |x_i(n)|^2 \quad (8)$$

Бұл теңдеу сигналдың деңгейін сипаттайды. Сипаттама белгілі бір классқа тиесілі, сигнал терезелері тізбегіндегі үлкен секірістерге ие болады. Мысалы, дыбыс үлкен көлемдегі кідірістерден және фонемалардан тұруы мүмкін.

Zero crossing rate (ZCR) сипаттамасын және оның теңдеуін әдебиеттерге шолу бөлімінде қарастырған болатынбыз. ZCR сипаттамасын шуыл өлемі ретінде қарастыруға болады. Сонымен қатар, белгілі нақтылықтағы сигналдың спектралдық сипаттамасын көрсете алады.

Энергия энтропиясы – сигналдың энергиялық деңгейіндегі өзгерістер өлшемі, қысқа уақыт ішіндегі энергия энтропиясы. Энергия энтропиясын анықтау үшін әрбір кадр бекітілген ұзақтықтағы  $K$  бөліктерге бөлінеді. Әрі қарай, әрбір  $j$  кадр бөлігіне  $E_{subframe_j}$  энергияны есептеп,  $E_{shortframe_i}$  жалпы энергияға бөлеміз. Кадр бөліктері үшін энергия мәнінің қорытынды тізбегі  $e_j$ ,  $j=1, \dots, K$  төмендегідей өрнектеледі:

$$e_j = \frac{E_{subframe_j}}{E_{shortframe_i}} \quad (9)$$

Мұндағы:

$$E_{shortframe_i} = \sum_{k=1}^K E_{subframe_k} \quad (10)$$

Әрі қарай  $e_j$  тізбегінің  $H(i)$  энтропиясы келесі теңдеуге (11) сәйкес анықталады:

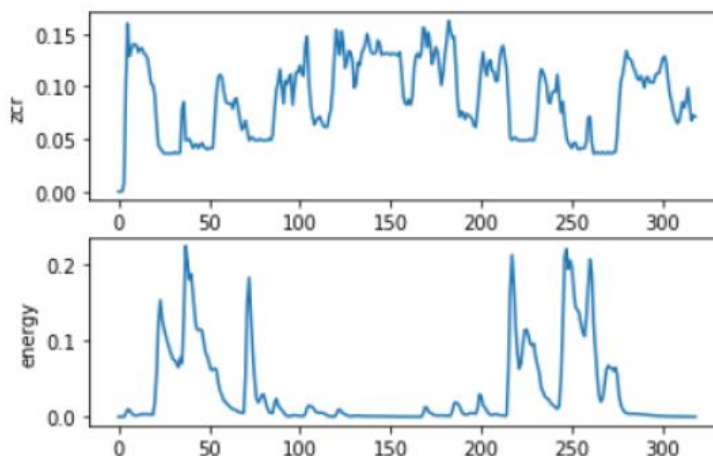
$$H(i) = -\sum_{j=1}^K e_j \cdot \log_2(e_j) \quad (11)$$

Энергияның кенеттен өзгеруі кезінде қорытынды нәтиже азаяды. Тиісінше, бұл сипаттама оқ атудың, жарылыстың және энергия мәнінің тез өзгеруі бар дыбыстардың басталуын анықтауға қолданылады.

Белгілерді алу үшін аудиосигнал алдымен қысқа сегменттерге немесе кадрларға бөлініп, әрбір кадрдан тиісті сипаттамалар алынады. Нәтижесінде қысқамерзімді сипаттамалық векторлар алынады. Сегменттердің өлшемі әр қалай болуы мүмкін, алайда, 20 мс-тан 100 мс-қа дейінгі ұзақтық жиі қолданылады.

Сипаттамалар тізбегі 50 мс кадр ұзақтығы және 25 мс қадам ұзындығымен алынды.

Сипаттамаларды алу үшін pyAudioAnalysis кітапханасы қолданылды. Алынған сипаттамалар 2-суретте көрсетілген.



2 – сурет – Кадрларға байланысты ZCR және энергияның мәндері

Сонымен қатар pyAudioAnalysis кітапханасы аудиосигналдың спектрограммасын да алуға мүмкіндік береді [15].

Кітапханадағы функция 68 сипаттаманы алды және аудиофайл 68x400 өлшеміндегі матрица түрінде ұсынылды. Белгілерді алу жылдамдығы аз. Мәліметтің 6500 үлгісін өңдеу үшін шамамен 2 сағат уақыт жұмсалды. Алайда бұл жылдамдық аудиосигналдарды нақты уақыт режимінде өңдеуге мүмкіндік беретіндігін ескерген жөн.

Қорытынды массив барлық аудиожазбалар үшін сипаттамалардан құралады және Pandas кітапханасындағы мәліметтер құрылымы Data Frame-де сақталады. Кестенің бірінші бағанында файлдың аты, екінші бағанда сипаттамалардың сандық мәндері, ал үшінші бағанда дұрыс класстардың векторлары сақталады. Онтология файлы вектор құру үшін қолданылады.

### КЛАССИФИКАЦИЯ ЖӘНЕ МӘЛІМЕТТЕРДІҢ КӨЛЕМІН АРТТЫРУ

Жұмыста конволюциялық желілердің кіші көлемін оқыту үшін үш әдіс қолданылды. Әдістер, өз кезегінде, алгоритмнің тиімділігін арттыруға және реттеу құралы ретінде әрекет етті. Әрі қарай салмақтар кіші dataset-пен жұмыс істеу үшін қолданылады.

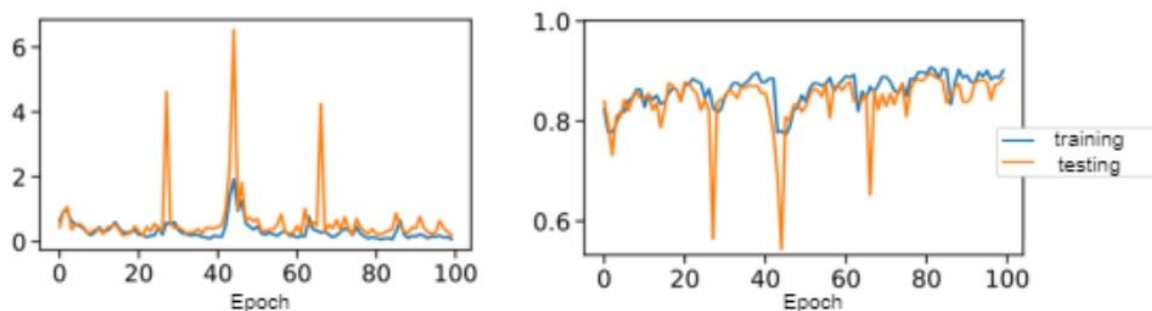
Аудио файлдарды түрлендірудің көптеген әдістері бар: кездейсоқ шуылдар қосу, уақыттық жылжытулар, жазба жылдамдығын және тонын өзгерту және т.б. Мәлімет көлемін арттырудың аталған әдістерін қолдану модель сапасының едәуір артуына ықпал етеді. Бұл алгоритмдерді әрбір аудио жазбаға қолдану арқылы, мәліметтердің көлемін екі есе арттыру мүмкін болады. Жұмыста аудиожазбалар шуыл қосу және дыбыс биіктігін арттыру арқылы түрлендірілді. Шуылдарды қосу үшін NumPy кітапханасы қолданылды [16]. Ол өз кезегінде кездейсоқ міндер тізбегін генерациялайды. Тізбек элементтері шуыл факторизациясы параметріне көбейтіледі. Шуыл факторизациясы – кездейсоқ шуылдың аудиожазбаға әсер ету күшін өзгертеді. Дыбыс биіктігін өзгерту үшін libros кітапханасындағы pitch\_shift функциясы қолданылды [17].

### НӘТИЖЕЛЕР

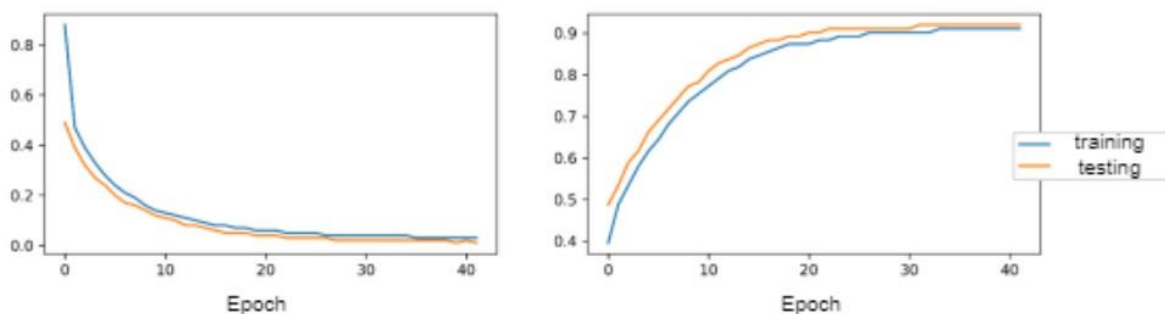
Dataset scikit-learn кітапханасына енгізілген стратификация функциясын қолдану арқылы бөлінді. Мәліметтердің 70% оқытуға, ал 30% тестілеуге қолданылды. Жоғалту функциясы (loss function) Adam алгоритмін қолдану арқылы оңтайландырылды. Бұл алгоритм қарапайым

стохастикалық градиентті түсудің алдын алуға арналған әдіс болып табылады. ReLU неғұрлым жоғары үйлесімділікті қамтамасыз етуі және градиентті өшулікке тәуелсіз болуы есебінен, активация функциясы ретінде қолданылды [18].

Қарапайым модель мен VGG моделі кем дегенде 40 қайталау арқылы тестіленді. Жоғалту функциясы арта бастаған кезде оқыту тоқтатылып отырды. Қарапайым модель 100 рет қайталай отырып оқытылды.



3 – Сурет – Бинарлы классификация үшін қарапайым модель бойынша алынған жоғалту функциясы мен нақтылығы



4 – сурет – Мультижіктеу үшін VGG модель бойынша алынған жоғалту функциясы мен нақтылығы

Бинарлы жіктеу жүргізу нәтижесінде, оқытуды 100 рет қайталаудан соң қарапайым модель 92,1% нақтылықты, ал мультижіктеу тапсырмасы 85,7% нақтылықты көрсетті (3-сурет). Ал нейрондық желі 40 қайталаудан соң бинарлы жіктеу үшін 98,8%, мульти жіктеу тапсырмасы 92,6% нақтылықты көрсетті (4-сурет).

4-кесте - Аудио оқиғаларды анықтау

Оқиға түрі	Нақтылығы (accuracy)	Precision	Recall	F1 score	AUC ROC
Қару дыбысы	0,9178	0,9245	0,9427	0,8945	0,9748
Терезе сынуы	0,9372	0,9765	0,9215	0,9154	0,9578
Өрт	0,9435	0,9346	0,9215	0,9345	0,9576
Сирена	0,9537	0,9462	0,9876	0,9642	0,9623
Жарылыс	0,8132	0,8254	0,8352	0,8124	0,9348
Жылау	0,8635	0,8524	0,8864	0,8754	0,9467
Иттің үруі	0,8456	0,8325	0,8571	0,8254	0,9425
Айқай	0,8654	0,8452	0,8576	0,8457	0,9472

## ҚОРЫТЫНДЫ

Жұмыс барысында қойылған тапсырмалар орындалды. Мәліметтерді енгізу және сипаттамаларды алу жүзеге асырылды. Мәліметтер жинағы көлемін ұлғайтудың шешімі табылды. Нейрондық желілер негізінде алынған модельдер дыбыстық оқиғалардың кең спектрін жіктеуге мүмкіндік береді. Қорытынды нәтиже алдыңғы жұмыстарда к жақын көршілер мен тірек векторлары әдістерімен салыстырғанда жоғары нәтижелер алынды.

Жұмыстың өзектілігі ретінде терең оқыту әдістерін қолдану арқылы, яғни нейрондық желілер негізінде бинарлы жіктеу үшін қарапайым модель және мультижіктеу тапсырмасы үшін VGG моделінің оқытылғандығын атап өткен жөн. Жұмыстың жаңалығы - өзіміз жеке жинақтаған data set негізінде файлдардың көлемін арттыру әдісін қолданып, түрлендіру арқылы модель нақтылығының жоғарғы мәнін алуда болып табылады.

Алдыңғы жұмыстарда VGG моделі 74% нақтылық көрсетсе, зерттеу жұмысы барысында қайта оқытылған VGG моделінің нақтылығы 85% дейін артты. Ескерте кететін жайт, VGG моделіндегі жасырын қабаттардың санын арттыру нақтылықты аздаған шамаға ғана арттырады. Сондықтан неғұрлым параметрлер саны аз қарапайым модельдерге назар аударған жөн. Себебі есептеу күрделілігінің артуы оқыту уақытының ұлғаюына әсер етеді.

Аудиожазбалар жіктеуі жылдамдығы нақты жұмыс режимінде жұмыс істеуге мүмкіндік бере алады. Зерттеудегі келесі қадам – кез-келген құрылғыдағы микрофон арқылы дыбысты автоматты түрде жазып, аудиоағынды сол мезетте сараптап, дыбыстық оқиғаны жіктеу.

Алынған жіктегіш түрлі салаларда қолданылуы мүмкін. Бақылау жүргізіп отырған мекен-жайдан төтенше оқиғалар туралы алдын-ала хабар беруге мүмкіндігі бар. Ерекше қорғауды талап ететін ғимараттардың, мұражай, әуежай, теміржол және авто бекеттер, банк, көрме, балабақша, мектеп, сауда, ойын-сауық орталықтары және басқа да қоғамдық орындардағы қауіпсіздікті қамтамасыз ету үшін қолданылады.

## ӘДЕБИЕТТЕР ТІЗІМІ/LIST OF REFERENCES

- [1] Delac, K., Grgic, M., & Grgic, S. Independent comparative study of PCA, ICA, and LDA on the FERET data set. *International Journal of Imaging Systems and Technology*, 2005, 15(5), 252–260.
- [2] Cao, L. J., Chua, K. S., Chong, W. K., Lee, H. P., & Gu, Q. M. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing*, 55(1-2), 321–336.
- [3] Rong, F. (2016). *Audio Classification Method Based on Machine Learning*. 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). Changsha, China. **Electronic ISBN:** 978-1-5090-6061-0
- [4] Pandiyan, “Mel-frequency cepstral coefficient analysis in speech recognition,” *Computing & Informatics* 2006, ICOCI’06, no. 2, pp. 2–6.
- [5] Jozef Vavrek, Eva Voz’arikov’a, Mat’u’s Pleva, and Jozef Juh’ar. Broadcast News Audio Classification using SVM Binary Trees. 2012 35th International Conference on Telecommunications and Signal Processing (TSP). Prague, Czech Republic. 3-4 July 2012. **Print ISBN:**978-1-4673-1117-5
- [6] M. Pleva, J. Juha’r, and A. C’iz’ma’r, “Slovak broadcast news speech corpus for automatic speech recognition,” in *Proc. of RTT ’07*, 2007.
- [7] L. Xie, Z.-H. Fu, W. Feng, and Y. Luo, “Pitch-density-based features and an SVM binary tree approach for multi-class audio classification in broadcast news,” *Multimedia Systems*, vol. 17, pp. 101–112, 2011.
- [8] M. Casey, “General sound classification and similarity in MPEG-7,” *Organised Sound*, vol. 06, pp. 153–164.
- [9] T. Sikora. *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. John Wiley & Sons, 2005.
- [10] Suliman, A., Omarov, B., Dosbayev, Zh. Detection of impulsive sounds in stream of audio signals. 2020 8th International Conference on Information Technology and Multimedia, ICIMU 2020, 2020, pp. 283–287, **DOI:** [10.1109/ICIMU49871.2020.9243540](https://doi.org/10.1109/ICIMU49871.2020.9243540)
- [11] Rajkumar Palaniappan, K. Sundaraj, «Respiratory Sound Classification using Cepstral Features and Support Vector Machine», 2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS). 978-1-4799-2178-2/13.

[12] Dosbayev, Zh., Abdrakhmanov, R., Akhmetova, O. and others. Audio Surveillance: Detection of Audio-Based Emergency Situations. ICCCI 2021: Advances in Computational Collective Intelligence 2021, 1463, pp. 413–424. DOI:10.1007/978-3-030-88113-9\_33

[13] Alsina-Pagès, R., Navarro, J., Alías, F., & Hervás, M. (2017). homeSound: Real-Time Audio Event Detection Based on High Performance Computing for Behaviour and Surveillance Remote Monitoring. *Sensors*, 17(4), 854. doi:10.3390/s17040854

[14] Adavanne, S., Politis, A., & Virtanen, T. (2018, July). Multichannel sound event detection using 3D convolutional neural networks for learning inter-channel features. In 2018 international joint conference on neural networks (IJCNN) (pp. 1-7). IEEE

[15] Derakhshan, M. O. R. A. D., Marvi, H. O. S. S. E. I. N., & Hassan Poor, H. (2019). Providing an Adaptive Model with two Adjustable Parameters for Audio Event Detection and Classification in Environmental Signals. *TABRIZ JOURNAL OF ELECTRICAL ENGINEERING*, 49(2), 565- 576.

[16] Arora, P., & Haeb-Umbach, R. (2017, October). A study on transfer learning for acoustic event detection in a real life scenario. In 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP) (pp. 1-6). IEEE.

[17] Lu, X., Shen, P., Li, S., Tsao, Y., & Kawai, H. (2018). Temporal Attentive Pooling for Acoustic Event Detection. In *Interspeech* (pp. 1354-1357).

[158] Wang, Y., Li, J., & Metze, F. (2019, May). A comparison of five multiple instance learning pooling functions for sound event detection with weak labeling. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 31-35). IEEE.

## CLASSIFICATION OF SOUND USING MACHINE LEARNING METHODS

Zh.M. Dosbayev<sup>1,2\*</sup>, B.S. Omarov<sup>2</sup>, B.M. Sadykova<sup>2</sup>, N.K. Smailov<sup>1</sup>

<sup>1</sup>Satbayev University, Almaty, Kazakhstan

<sup>2</sup>Kazakh National University named after Al-Farabi, Almaty, Kazakhstan

E-mail: [Zh.dosbayev@satbayev.university](mailto:Zh.dosbayev@satbayev.university), [d\\_jandos\\_93@mail.ru](mailto:d_jandos_93@mail.ru), [batyahan@gmail.com](mailto:batyahan@gmail.com),  
[mukhtarkyzy.bibi@gmail.com](mailto:mukhtarkyzy.bibi@gmail.com), [nur\\_aly.kz@mail.ru](mailto:nur_aly.kz@mail.ru)

**Abstract.** The purpose of the research work is to increase the accuracy of classification of the trained model using machine learning techniques used in the audio classification of data received from audio sensors. Various stages of audio data classification were separately discussed; an analysis was carried out based on the decisions of other authors. Methods of feature extraction, advantages of obtaining features, equations and applications were determined. Based on the obtained features, models were trained using machine learning algorithms, architectures were described, and testing results were described. The features, general characteristics, and classes of the used audio data set were also considered. The features of audio data preprocessing, filtering, and structural units of audio data are named. In previous research papers, we trained the model using shallow learning algorithms, the relevance of this work is the training of a simple model for binary classification and a VGG model for setting multiclassification were carried out using deep learning methods, based on neural networks. As a result of the study, the accuracy of binary classification using a neural network was 98.8% and the accuracy of 92.6% for the multiclassification task. And the accuracy of the VGG model has increased from 74% to 85%. The use of high-precision models trained by machine learning methods allows you to ensure safety in public places and indoor environments. As the novelty of the work can be noted the specifics of the methods used to transform dataset (data augmentation) and audio files to increase efficiency, the variety of features and algorithms obtained, as well as the achievement of the highest accuracy values. The directions requiring further research were also taken into account.

**Keyword:** audio classification, machine learning algorithms, audio Analytics, neural network, feature extraction.

## КЛАССИФИКАЦИЯ ЗВУКА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Ж.М. Досбаев<sup>1,2 \*</sup>, Б.С. Омаров<sup>2</sup>, Б.М. Садыкова<sup>2</sup>, Н.К. Смайлов<sup>1</sup>

<sup>1</sup>Сатпаев университет, Алматы, Казахстан

<sup>2</sup>Казахский Национальный университет имени аль-Фараби, Алматы, Казахстан

E-mail: [Zh.dosbayev@satbayev.university](mailto:Zh.dosbayev@satbayev.university), [d\\_jandos\\_93@mail.ru](mailto:d_jandos_93@mail.ru), [batyahan@gmail.com](mailto:batyahan@gmail.com),  
[mukhtarkyzy.bibi@gmail.com](mailto:mukhtarkyzy.bibi@gmail.com), [nur\\_aly.kz@mail.ru](mailto:nur_aly.kz@mail.ru)

**Аннотация.** Целью исследовательской работы является повышение точности классификации изучаемой модели с применением методов машинного обучения, используемых при проведении аудио классификации данных, полученных от аудиосенсоров. Отдельно были затронуты различные этапы классификации аудиоданных, проведен анализ по решениям других авторов. Были определены методы, особенности и преимущества получения признаков, уравнения и применение. По полученным признакам были проведены обучения моделей с использованием алгоритмов машинного обучения, описаны архитектуры и результаты тестирования. Также были рассмотрены особенности, общие характеристики, классы используемых аудио dataset. Названы особенности предварительной обработки аудиоданных, фильтрации, структурные единицы аудиоданных. В предыдущих исследовательских работах нами проводилось обучение модели с помощью традиционных алгоритмов обучения, актуальностью данной работы является обучение простой модели для бинарной классификации и модели VGG для задания мультиклассификаций проводились с использованием методов глубокого обучения, то есть на основе нейронных сетей. В результате исследования была получена точность бинарной классификации с использованием нейронной сети 98,8% и точность 92,6% по задаче мультиклассификаций. А точность модели VGG увеличилась с 74% до 85%. Применение моделей высокой точности, обученные методами машинного обучения, позволяет обеспечить безопасность в общественных местах и внутри зданий. В качестве новизны работы можно отметить специфику используемых методов преобразования dataset и аудиофайлов для повышения эффективности, разнообразие полученных признаков и алгоритмов, а также достижение наибольших значений точности. Были также учтены направления, требующие дальнейшего исследования.

**Ключевые слова:** классификация звука, алгоритмы машинного обучения, аудиоаналитика, нейронные сети, извлечение признаков.



## Басылымның шығыс деректері

Мерзімді баспасөз басылымының атауы	«Алматы энергетика және байланыс университетінің Хабаршысы» ғылыми-техникалық журналы
Мерзімді баспасөз басылымының меншік иесі	«Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті» коммерциялық емес акционерлік қоғамы, Алматы, Қазақстан
Бас редактор	Профессор, т.ғ.к., В.В. Стояк
Қайта есепке қою туралы куәліктің нөмірі мен күні және берген органның атауы	№ KZ14VPY00024997, күні 17.07.2020, Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігі
Мерзімділігі	Жылына 4 рет (тоқсан сайын)
Мерзімді баспасөз басылымының реттік нөмірі және жарыққа шыққан күні	Жалпы нөмір 57, 2-басылым, 2022 жылғы 30 маусым
Басылым индексі	74108
Басылым таралымы	200 дана
Баға	Келісілген
Баспахана атауы, оның мекен-жайы	«Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті» КЕАҚ баспаханасы, Байтұрсынұлы көшесі, 126/1 үй, А120 каб.
Редакцияның мекен-жайы	050013, Алматы қ., «Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті» КЕАҚ, Байтұрсынұлы к-сі, 126/1 үй, каб. А 224, тел.: 8 (727) 292 58 48, 708 880 77 99, e-mail: <a href="mailto:vestnik@aes.kz">vestnik@aes.kz</a>

## Выходные данные

Название периодического печатного издания	Научно-технический журнал «Вестник Алматинского университета энергетики и связи»
Собственник периодического печатного издания	Некоммерческое акционерное общество «Алматинский университет энергетики и связи имени Гумарбека Даукеева», Алматы, Казахстан
Главный редактор	Профессор, к.т.н., Стояк В.В.
Номер и дата свидетельства о постановке на учет и наименование выдавшего органа	№ KZ14VPY00024997 от 17.07.2020 Министерство информации и общественного развития Республики Казахстан
Периодичность	4 раза в год (ежеквартально)
Порядковый номер и дата выхода в свет периодического печатного издания	Валовый номер 57, выпуск 2, 30 июнь 2022
Подписной индекс	74108
Тираж выпуска	200 экз.
Цена	Договорная
Наименование типографии, ее адрес	Типография НАО «Алматинский университет энергетики и связи имени Гумарбека Даукеева», ул. Байтұрсынұлы, дом 126/1, каб. А 120
Адрес редакции	050013, г. Алматы, НАО «Алматинский университет энергетики и связи имени Гумарбека Даукеева», ул. Байтұрсынұлы, дом 126/1, каб. А 224, тел.: 8 (727) 292 58 48, 708 880 77 99, e-mail: <a href="mailto:vestnik@aes.kz">vestnik@aes.kz</a>

## Issue output

Name of the periodical printed publication	Scientific and technical journal "Bulletin of the Almaty University of Power Engineering and Telecommunications"
Owner of the periodical printed publication	Non-profit joint-stock company "Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeyev", Almaty, Kazakhstan
Chief Editor	Professor, candidate of technical sciences Stoyak V.V.
Number and date of the registration certificate and the name of the issuing authority	№ KZ14VPY00024997 from 17.07.2020 Ministry of Information and Social Development of the Republic of Kazakhstan
Periodicity	4 times a year (quarterly)
Serial number and date of publication of a periodical printed publication	Number 57, edition 2, June 30, 2022
Subscription index	74108
Circulation of the issue	200 copies
Price	Negotiable
The name of the printing house, its address	Printing house of Non-profit joint-stock company "Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeyev", 126/1 Baitursynuly str., office A 120, Almaty, Republic of Kazakhstan
Editorial office address	050013, Non-profit joint-stock company "Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeyev", A 224, tel.: 8 (727) 292 58 48, 708 880 77 99, e-mail: <a href="mailto:vestnik@aes.kz">vestnik@aes.kz</a>