



VPP
Valsts pētījumu
programma



Projekts Nr. VPP-COVID-2020/1-0025

**Jaunās tehnoloģijas Covid-19 pacientu tēmētai monitorēšanai, testēšanai un
terapijai (3-T Project)**

Projekta papildu (specifiskie) rezultāti Nr.17

Jauna monitorēšanas sistēma

Tehnoloģijas apraksts

2. pielikums

Autori:

J. Teļičko, D. Vidulejs, A. Jakovičs

Rīgā, 2020. gada decembrī

Saturs

Ievads	3
1. Monitorēšanas sistēma	3
2. Tiešie sensoru mērījumi	3
3. Cilvēku videoskaitītājs	6
4. Riska indikatortornis	11
5. Skaņas sensors	11
6. Kopsavilkums	14
7. Atsauces	15



Pārvaldības sistēmu un programmatūras uzlabošana

Ievads

Monitorēšanas sistēmas ļauj sasniegt augstāko komfortu ēkās, taču parasti to noteiktie parametri nav pietiekami, lai analizētu epidemioloģisko apdraudējumu. Pandēmijas un pieaugošās saslimstības dēļ radās nepieciešamība pēc monitorēšanas sistēmām, kas spētu sniegt nepieciešamo informāciju infekcijas risku analīzei. Laicīgi informējot cilvēkus par riskiem, šāda sistēma varētu ne tikai palielināt drošību ēkās, bet arī ietaupīt medicīnas personāla darba resursus.

Monitorēšanas sistēma

Monitorēšanas sistēmu veido vairāki sensori (oglekļa dioksīda, mitruma, putekļu), kā arī kamera un kapacitatīvs mikrofons ar papildus iebūvētu mākslīgā intelekta datu apstrādi. Šādu komplektu ir viegli izvietot. Sistēmas uzdevums ir sniegt nepieciešamo informāciju, lai novērtētu infekcijas risku konkrētā telpā. Sakaru protokoli un komponentes ir izvēlētas tā, lai sistēma būtu viegli mērogojama un saglabātu zemas izmaksas.

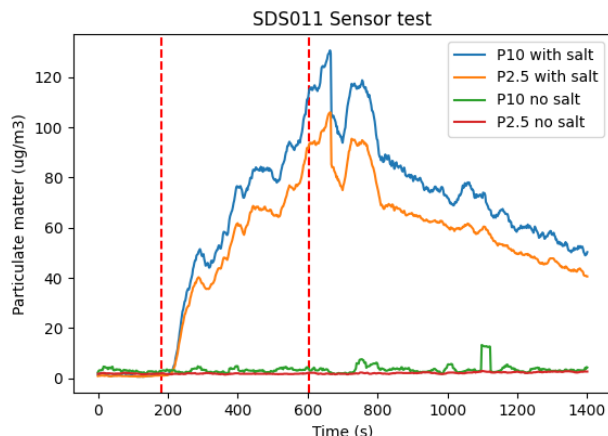
Par projekta pamatu tika izvēlēts *Jetson Nano* modulis tā zemo izmaksu un iebūvētā grafiskā procesora (*GPU*) dēļ. Pirmkārt, šādi integrētie datori ļauj apstrādāt datus iebūvētos neironīklos (*Neural Networks - NN*), kā arī samazina privāto datu aizsardzības sarežģītību, jo dati nav jāuzglabā. Tajā pašā laikā *Jetson Nano* varētu darboties kā starpnieks, lai uzturētu sakarus starp citiem sistēmas moduļiem. Tā kā *Jetson Nano* ir *Linux* operētājsistēma, *ssh* varētu izmantot attālās konfigurācijas un atklādošanas sistēmai.

Tiešie sensoru mērījumi

CO₂, mitruma un putekļu mērījumus var ietekmēt vairāki faktori, piemēram, tuvumā esošie sildītāji un ventilācijas kanāli. Šī iemesla dēļ tika nolemts izveidot atsevišķu plati, jo pastāv liela varbūtība, ka optimālā kameras uzstādīšanas vieta varētu būt ventilācijas kanāla vai cita objekta tuvumā, kas varētu ietekmēt mērījumus.

Tika veikti sākotnējie pētījumi un vairāki eksperimenti ar mērķi noteikt, kuri mērījumu parametri varētu sniegt vairāk informācijas par telpu stāvokli. Viens no neparastiem mērījumu parametriem ir putekļu sensors. Ideja ir izmērīt cietās daļiņas, kas var saglabāties no cilvēka izelpotā aerosola. Lai apstiprinātu šo teoriju, tika veikts eksperiments, kura laikā pjezoelektriskās ierīces darbinātā gaisa mitrinātājā ielēja destilētu ūdeni un 5% sāls šķīdumu. PM₁₀ un PM_{2.5} tika mērīti trīs dažādos stāvokļos - kad mitrinātājs bija izslēgts, darbojoties ar destilētu ūdeni un darbojoties ar 5% sāls šķīdumu. Vērtības tika noteiktas ar SDS011 sensoru 6 x 7 m lielā telpā aptuveni 7 m attālumā no mitrinātāja.

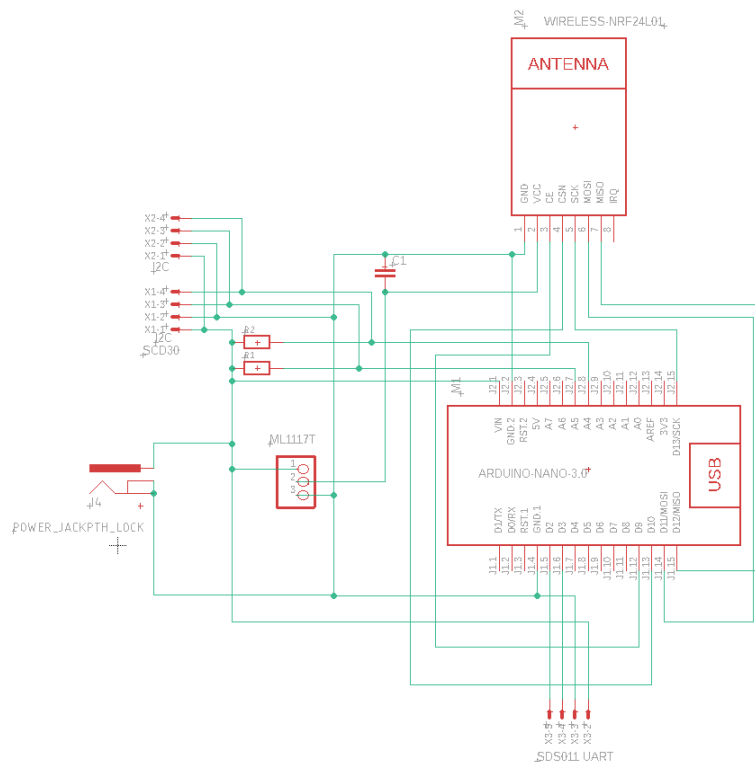
Kad mitrinātājs bija izslēgts, PM vērtības bija tādas pašas, kā destilētam ūdenim (PM₁₀ = 10, PM_{2,5} = 7). Kad mitrinātājā ielēja 5% sāls šķīdumu, ieslēdzot mitrinātāju, vērtības sāka ievērojami pieaugt un aptuveni 400 sekundēs pieauga par vairāk nekā 100 vienībām katrai vērtībai - skatīt tālāk redzamo grafiku:



1. att. PM mērijumi ar gaisa mitrinātāju, kas uzpildīts ar sāls šķīdumu vai destilētu ūdeni. Sarkanais kvadrāts atzīmē datus, kas iegūti, kad gaisa mitrinātājs bija ieslēgts.

Eksperimenta rezultātā tika nolemts mūsu sistēmā iekļaut PM sensoru.

CO2 mērijumiem tika izvēlēts SCD30 sensors tā stabilo mērijumu rezultātu dēļ. Šie sensori sazinās caur I2C protokolu, un tos var darbināt ar 5 V, tāpat kā SDS011. Turklāt minētajam SCD30 ir iebūvēti relatīvā mitruma un temperatūras sensori. SCD30 un SDS011 trūkums ir tas, ka abiem nepieciešama diezgan liela strāva - attiecīgi 20 mA un 70 mA. Kā barošanas avots tika izvēlēta elektroapgāde, jo piemērota akumulatora iekļaušana būtu dārga, turklāt tam būtu lieli izmēri. Datu pārraidei tika izvēlēts nrf24 radiomodulis, jo viena moduļa cena ir zem 2 USD, katrā 125 frekvenču kanālā ir 6 datu izvadi, kas, ievērojot pareizu sensoru topoloģiju, varētu ļaut datu pārraidi no vairāk nekā 700 datu punktiem. Iemesls, kāpēc netika izmantoti tādi risinājumi kā esp8266 vai esp32 ar Wi-Fi un viegli pieejamiem MQTT protokola piemēriem, ir Ethernet savienojuma izveides sarežģītība drošības apsvērumu dēļ. Piemēram, lai saņemtu piekļuvi internetam Latvijas Universitātē, jābūt reģistrētam ierīces MAC un Wi-Fi savienojumam ir nepieciešama pierakstīšanās caur tīmekļa lietojumprogrammu, kas dažādās vietās varētu atšķirties un kam būtu nepieciešama atšķirīga savienojuma izveides procedūra. Par sensoru plates pamatu tika izvēlēts *Arduino Nano* ar *Atmega328P* mikrokontroleri. Šī ir ļoti kompakta plate ar I2C, SPI un UART sakaru protokolu, kas nepieciešami saziņai ar citām komponentēm. Saskaņā ar mūsu mērijumiem nrf24 ar jaudas pastiprinātāju pārraides sākumā varētu būt nepieciešami vairāk nekā 100 mA, tāpēc ķēdei tika pievienots papildu 3,3 V stabilizators. Izveidotā shēma ir parādīta 2. attēlā.



2. att. Shēma

Lai aizsargātu un fiksētu komponentes, tika izdrukāts plastmasas korpuss.



3. att. Korpūsā saliktā plate

Eksperimentiem tika izvēlētas divas telpas Latvijas Universitātē - 508. telpa, ko izmantoja laboratorijas uzdevumiem, un 523. telpa, kas ir tipisks birojs ar 8 darba vietām. Abās telpās sensori tika uzstādīti katras telpas aptuvenajā centrā 30 cm attālumā no griestiem. Skat. 4. att.



4. att. Sensora lokācijas piemērs reālos apstākļos

Kā jau minēts iepriekš, datu pārsūtīšanai uz *Jetson Nano* bezvadu režīmā tika izmantots nrf24 radiomodulis. *Jetson Nano* ir savienots ar Ethernet un, izmantojot MQTT protokolu, retranslē datus uz riska skaitļošanas vienību.

Cilvēku videoskaitītājs

Daudzos lielveikalos cilvēku skaitīšanai izmanto optiskos pārtraukuma sensorus (*optical interrupt sensors*). Šādu sistēmu trūkums ir tas, ka gadījumos, kad uzstādīts tikai viens sensors, nav iespējams noteikt precīzu cilvēku skaitu - cik cilvēku ir ienākuši un cik izgājuši. Papildinot šādu sistēmu ar otro sensoru, lai noteiktu personas trajektoriju, teorētiski ir iespējams iegūt precīzu cilvēku skaitu, tomēr, ja cilvēki vienlaikus ienāk un iziet vai pārvietojas grupās, ir ļoti augsta kļūdas iespēja un sistēmai būs nepieciešama atkārtota kalibrēšana.

Alternatīvs risinājums ir kamera un attēlu papildapstrāde pēc mākslīgā intelekta algoritma. Lai neironu tīkli (NN) darbotos pareizi, ir nepieciešams izveidot arhitektūru un apmācīt tīklu, izmantojot vairākus piemērus. Tas rada vairākus sarežģījumus, jo šajā gadījumā fotoattēlu saglabāšanai birokrātiski apgrūtinājumi, kas saistīti ar personas datu apstrādi, turklāt nepieciešamā datu daudzuma iegūšana un marķēšana var būt laikietilpīga. Lai laicīgi īstenotu projektu, sensora izstrādē tika izmantotas gatavas arhitektūras un svērumi ar attēla iepriekšēju apstrādi.

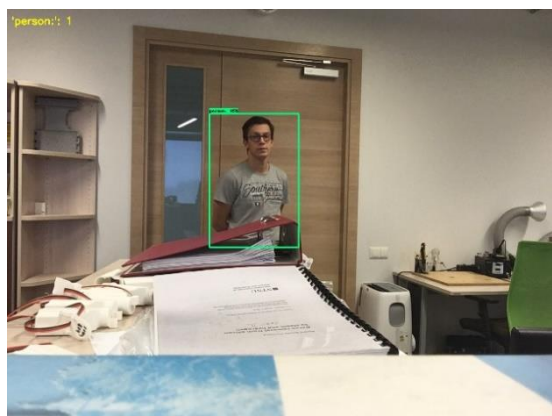
Jetson Nano tika izvēlēts ne tikai zemās cenas un *GPU* dēļ, bet arī tāpēc, ka operētājsistēmas programmatūrā jau ir iekļauta *CUDA* pakotne un ir iespējama tādu bibliotēku instalēšana kā *TensorFlow* pilnā versija. Testēšanas laikā tika atklāts plates darbības traucējums pēc strāvas pieslēgšanas, proti, ja netiek izmantota papildu dzesēšana, tiek aktivizēta datu plūsmas ierobežošana un dramatiski krītas veiktspēja.

Sākotnējo testu laikā tīmekļa kamera tika pieslēgta *Jetson Nano* caur USB. Skat. 5. att.



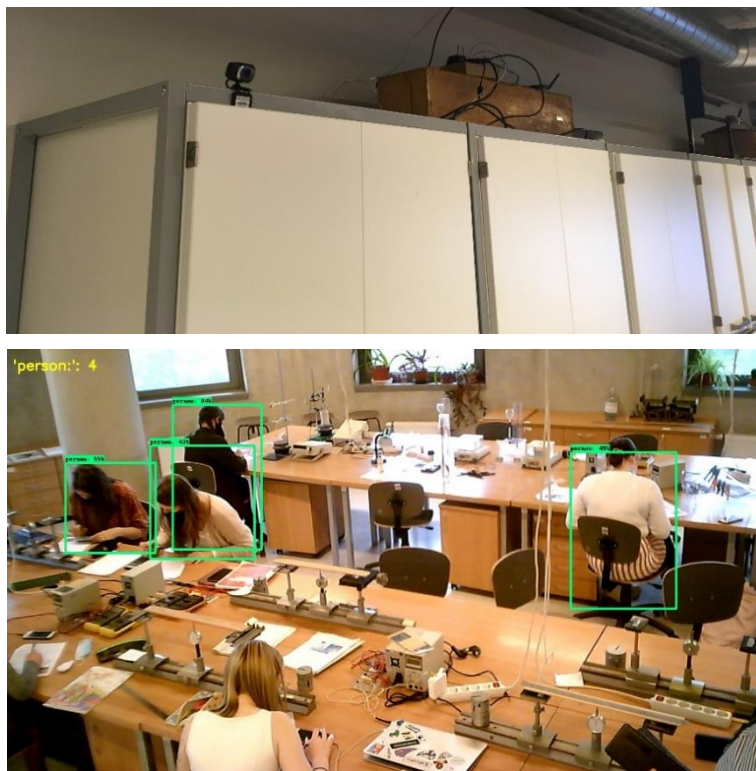
5. att. Jetson Nano ar tīmekļa kameru

Objektu detektēšanā un semantiskā segmentēšanā bieži izmanto konvolucionālos neironu tīklus jeb neirotīklus (*convolutional neural networks*) [1, 2, 3]. Sākotnējiem testiem tika izvēlēta atklātā pirmkoda programmatūra - *TensorFlow Object Detection API* ar "SSD with Mobilenet" modeli, ko nodrošināja Ahmet Özlü [4]. Kad kamera tika novietota uz biroja galda, tā viegli varēja atpazīt cilvēkus, kas gāja garām. Skat. 6. att.



6. att. Cilvēka atpazīšanas piemērs

Tad kamera tika pārvietota uz 508. laboratorijas telpu, kur tā tika novietota gandrīz optimālā stāvoklī, lai novērotu visu telpu. Skat. 7. att.



7. att. 508. telpa tīmekļa kamerā

Kā redzams fotoattēlā, kameras redzamības lauks nav pietiekams, lai attēlotu visu telpu, un cilvēki, kas atrodas tuvu attēla robežām, netiek atpazīti, visticamāk, tieši konvolūcijas dēļ. Tāpēc, lai atainotu visu telpu, tika izvēlēta kamera ar 170° leņķi (IMX219-170). Priekšrocība ir arī tāda, ka šis modelis ir daudz mazāks, tam ir ievērojami labāka izšķirtspēja līdz 3280×2464 un ar MIPI CSI-2 savienojumu cena ir zemāka nekā USB tīmekļa kamerām, kuru cena pandēmijas dēļ ir ievērojami augusi. Jaunā kamera sniedz visas telpas pārskatu, bet, kā redzams 8. attēlā, attēlam ir *fisheye* efekts.



8. att. IMX219-170 kamera, oriģinālais attēls

Tā kā modeļa apmācības datu kopai nebija *fisheye* efekta, pasliktinājās neironu tīkla precizitāte.



Lai atrisinātu šo problēmu, tika izmantots *OpenCV* bibliotēkas *fisheye* korekcijas [5] aprēķināšanas K un D koeficients, vairākas fotogrāfijas tika uzņemtas ar melnbalto kalibrēšanas režģi. Mūsu gadījumā koeficienti bija šādi:

$DIM=(1920, 1080)$

$K=np.array([[466.4955119733443, 0.0, 986.3542823807117], [0.0, 348.20536430394856, 493.33460518144153], [0.0, 0.0, 1.0]])$

$D=np.array([[0.16707137577986436], [0.02625086444431634], [-0.05043318932800496], [0.008872094570022155]])$

1. tabula. K un D koeficienti IMX219-170 kameras modulim

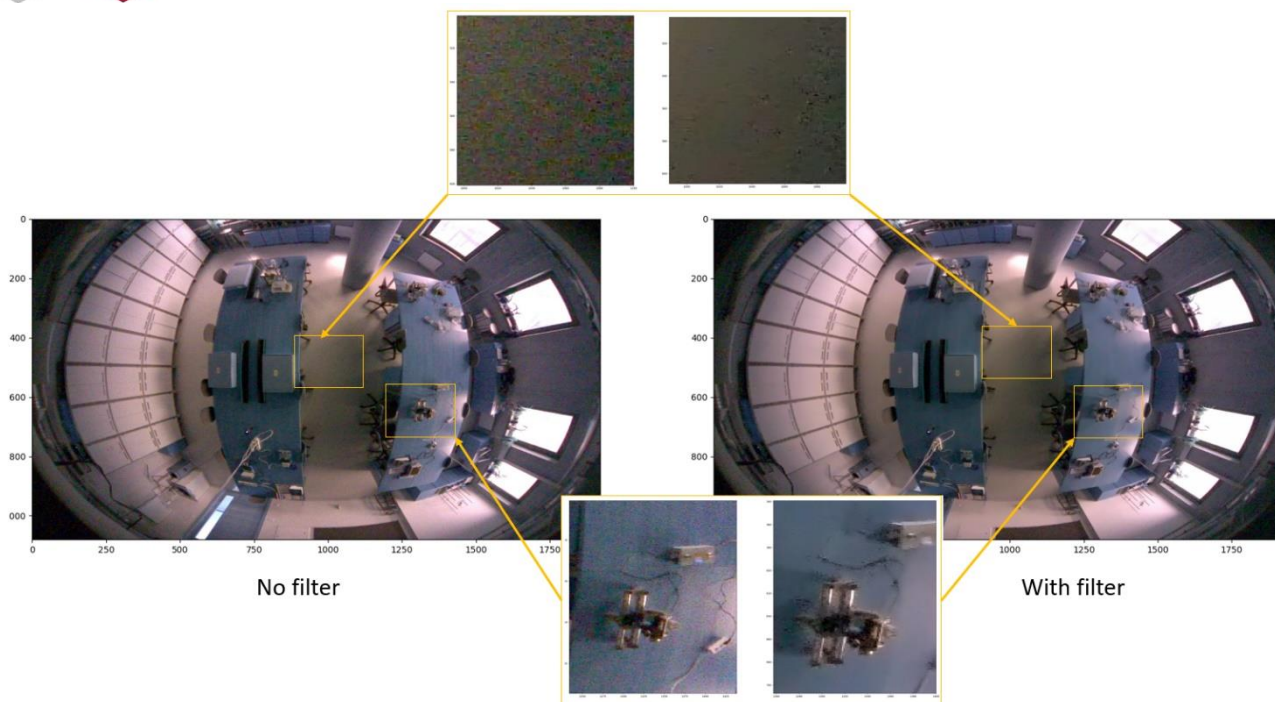
Rezultāts ir redzams 9. att.



10. att. *Fisheye* korekcijas rezultāti

Atpazīšanas precizitāte uzlabojās, taču sistēma slikti atpazīna cilvēkus pie tālākā galda. Šī iemesla dēļ tika pārbaudīti vairāki citi NN modeļi, piemēram, *YOLO v3*, *CSRNet* un *Mask R-CNN* [6, 7, 8]. Vislabākie rezultāti bija *Mask R-CNN* ar *mask_rcnn_coco* koeficientu. NN tika izvēlēts visiem tālākajiem eksperimentiem. Diemžēl pat ievērojami sarežģītākajai sistēmai bija grūtības atpazīt cilvēkus, kas atrodas tālu no kameras. Ir zināms, ka datorredzes zinātnei mazu objektu atpazīšana ir grūts uzdevums [9]. Tika pieņemts lēmums telpas vidū uzstādīt kameru ar skata leņķi no augšas uz leju; šādā veidā tiek samazināts maksimālais attālums līdz kamerai.

Kad naktīs un no rīta tika izslēgts apgaismojums, sistēma bieži vien pieļāva kļūdas, norādot, ka telpā ir vairāki cilvēki. Lai novērstu kļūdu, tika pievienota papildu pārbaudes funkcija un filtrs. Vispirms kamera aprēķina pikselu vidējo vērtību un pārbauda, vai tā ir augstāka par sliekšni - šādā veidā kamera identificē, vai attēls ir pilnīgi tumšs; ja tā, tad cilvēku skaits ir 0. Grūtāk ir no rīta, kad apgaismojums netiek ieslēgts, bet telpās varētu būt cilvēki. Pēc detalizētas vājā apgaismojumā uzņemtu attēlu izpētes kļuva skaidrs, ka šādos apstākļos kamera rada krāsu artefaktus. Tāpēc tika noteikts vēl viens sliekšnis Gaussian noise artefakta noņemšanai. Šim mērķim tika izmantots *fastNlMeansDenoisingColored* no *OpenCV*. Atšķirība



11. att. *Gaussian noise* artefakta filtrēšanas rezultāti

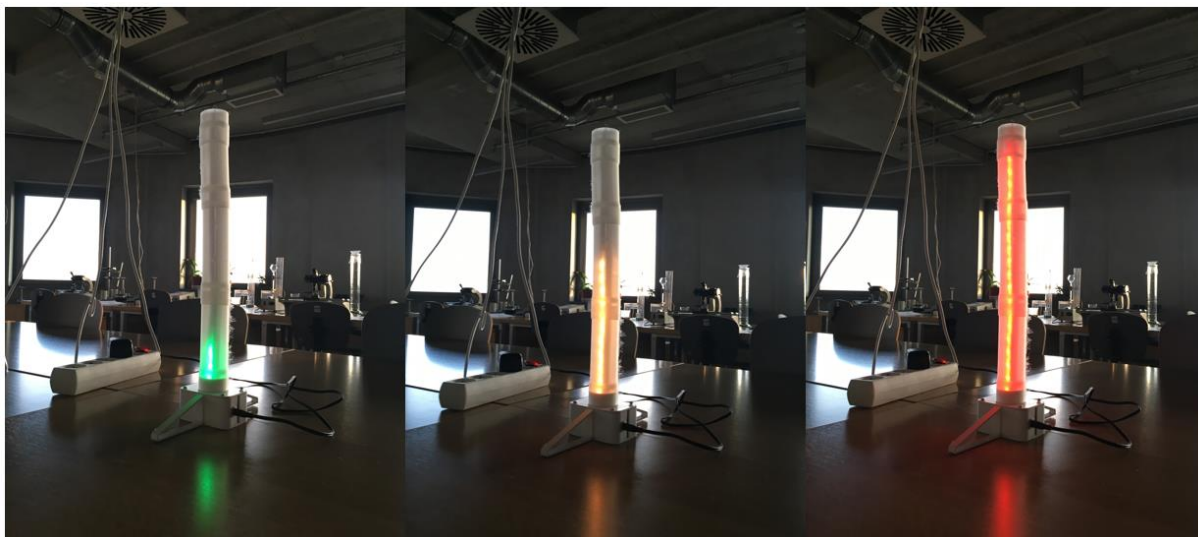


12. att. *Gaussian noise* artefakta filtrēšanas piemērs reāliem datiem

Tika novērots, ka NN gandrīz nekad nedeва kļūdaini pozitīvu rezultātu, bet samērā bieži norādīja mazāku cilvēku skaitu par patieso. Tāpēc tika pievienota 20 sekundes liela papildu buferatmiņa, kas tika aizpildīta pēc FIFO principa, nosakot NN atbildi pēc buferatmiņas maksimālā daudzuma. Pēc visiem pielāgojumiem NN spēja veiksmīgi saskaitīt cilvēkus birojā, bet diemžēl, ja cilvēki lielā laboratorijas telpā atradās tuvu viens otram, NN sniedza nepareizas vērtības, norādot mazāk cilvēku nekā faktiski novērots.

Riska indikatorornis

Tika uzstādīts LED tornis, lai informētu cilvēkus par potenciālo infekcijas risku. Kā pamats tika izvēlēts esp32 ar MQTT protokolu (gadījumā, ja nav pieejams bezvadu internets, kā alternatīvu var izmantot BLE ar nrf24) – šīs plātes abonē riska rādītāju un, saņemot vērtību, vizualizē signālu ar krāsu un konkrētu aktivēto diožu skaitu. Kā gaismas avots tiek izmantota programmējama LED virkne; komunikācija ar katru diodi tiek realizēta caur viena vadu protokolu. Tā kā esp32 loģikas līmenis ir 3,3 V un LED virkne darbojas ar 5 V, ir nepieciešams papildu loģikas līmeņa pārveidotājs.



13. att. Riska indikators

Skaņas sensors

Vērtējot pieejas cilvēku vokalizāciju atpazīšanai, ir veikti vairāki izšķiroši soļi un aprēķini, kas balstīti pašreizējā zinātniskajā literatūrā par mākslīgo intelektu un digitālo skaņas apstrādi, vadot tālāku nozares attīstību.

Pētījuma sākotnējie posmi bija vērsti uz atbilstošas datubāzes uzkrāšanu un nepieciešamo rīku izveidi, lai savienotu datus ar mākslīgo intelektu. Ir ārkārtīgi svarīgi modificēt datus tā, lai tie līdzinātos reālajām dzīves situācijām un statistiskajiem sadalījumiem, lai modelis izturētos līdzīgi arī ārpus testa apstākļiem. Tādēļ lielas pūles ir veltītas zinātniski motivēta skaņas specifisko datu augmentācijas moduļa izstrādei, kas paredzēts, lai paaugstinātu tajā apmācīto AI modeļu noturību un vispārējo precizitāti.

Datu augmentācijas modulis uzlabo datu kopas kvalitāti, ieviešot datu kopā nejaušus traucējumus, piemēram, troksni, kas veicina precīzu skaņu identifikāciju AI modelī, tādējādi uzlabojot spēju atšķirt trokšņus un veicinot modeļa precizitāti. Tiek izmantotas arī vairākas citas metodes, lai nodrošinātu, ka modelis iemācās atpazīt ne-nejaušas pazīmes, piemēram, piķu nobīdi (*pitch shifting*), laika nobīdi (*time shifting*), uz amplitūdu balstītu segmentu atkārtošanos [10] un trokšņa samazināšanu (*noise-trimming*) skaņām, kuru dati tiek apstrādāti. Pēc tam datu kopu var paplašināt vairākas reizes, nepalielinot redundanci.



Klase	Ilgums minūtēs
Klepus	67
Šķaudīšana	47
Smiekli	26
Sarunas	179
Troksnis	296

Kopējais audioparaugu ilgums katrā klasē pirms augmentācijas piemērošanas

Datu kopa ir sastādīta no vairākiem avotiem, proti, *ESC-50: Dataset for Environmental Sound Classification*, *VoxCeleb* datu kopa, *Speech Accent Archive*, *Freesound One-Shot Percussive Sounds* datu kopa, lietotāju iesniegtas skaņas *Freesound.org* un *Covid-19 Sounds App* no Kembridžas Universitātes. Mēs vēlamies pateikties Kembridžas Universitātes zinātniekiem par dalīšanos ar savu kvalitatīvo klepus skaņu datu kopu. Ir būtiski pieminēt, ka datu kopas sastādīšana ir viens no vissvarīgākajiem un izaicinošākajiem uzdevumiem AI modeļa izveidē. Nenovēršama ir saistīto problēmu rašanās. Pirmkārt, dati pēc savas būtības ir ļoti trokšņaini. Otrkārt, pašlaik to daudzums ir ļoti mazs. Covid-19 pandēmijas ietekmē zinātnieki un iestādes tiešsaistē sāk apkopot brīvprātīgo balsis ierakstus.

Vienlaikus ar datu kopas izveidi tika pētīts un izstrādāts AI modelis, izvēloties īpašus arhitektūras dizainus, kas saderīgi ar integretā datora skaitļošanas resursiem. Modelis ir balstīts uz atvieglotās sekvenču apguves algoritmu (*lightweight sequence learning algorithm*), īpaši *Bidirectional Long-Short-Term-Memory (BLSTM)* - rekurentā neironu tīkla (*Recurrent Neural Network*) veidu, kuru baro *Mel-Frequency Cepstral* koeficienta (MFCC) funkcija. AI arhitektūra spēj uztvert īslaicīgo datu (*temporal data*) nianšes neatkarīgi no to ilguma (*temporal duration*), un projektā izvēlēta funkcija pārvalda cilvēka plašā dzirdes spektra segmentāciju.

Pētnieki arvien plašāk izmanto konvolucionālo neironu tīklu iespējas uz skaņu balstītu uzdevumu risināšanā. Veiksmīga CNN tīkla piemērs ir milzīgais *Google Audioset* modelis, kurā klasificētas vairāk nekā 30 000 dažādas pazīmes un apmācīti 5,4 miljoni audiostundu [11]. Negatīvie aspekti ir augstā CNN resursietilpība, tāpēc tika pieņemts lēmums izmantot vieglāku arhitektūru. Mūsdienīgie augsta līmeņa rezultāti “Anomalous Sound Detection for Audio Surveillance” pētījumā [12] iedrošināja adaptēt RNN arhitektūru. Turklāt anomālu skaņu noteikšanas problēma ļoti līdzinās klepus un šķavu noteikšanas problēmai, jo šie audionotikumi ir gan reti, gan īsi. Modelis ir monofonisks. Tas nozīmē, ka tas ir apmācīts atpazīt tikai vienu - izteiktāko skaņu paraugā, jo, izvēloties polifonisku modeli, tiek nevajadzīgi palielināta modeļa sarežģītība.

**VPP**

Valsts pētījumu programma

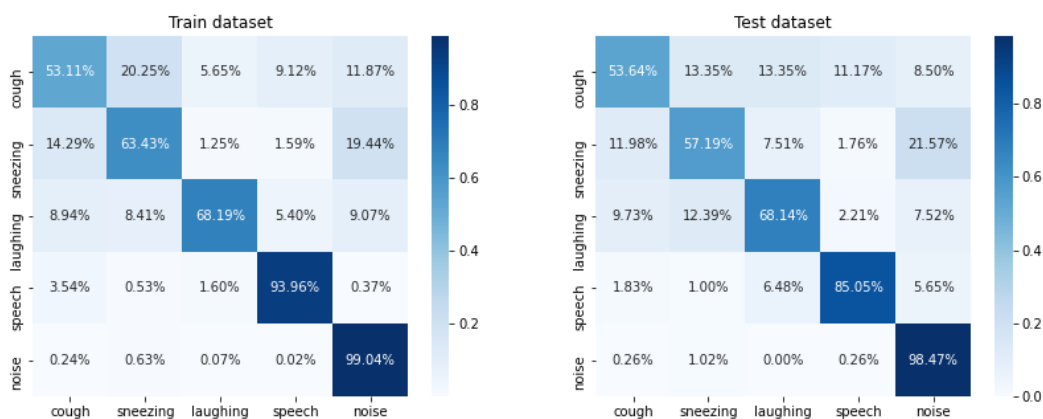
3T CONCEPT
TRACING · TESTING · TREATMENT

“Jaunās tehnoloģijas Covid-19 pacientu tēmētai monitorēšanai, testēšanai un terapijai (3-T Project)” (Nr. VPP-COVID-2020/1-0025)

Layer (type)	Output Shape	Param #
batch_normalization_1 (Batch Normalization)	(None, 99, 12)	48
LSTM_80_Neurons (Bidirectional LSTM)	(None, 99, 160)	59520
gaussian_noise_1 (Gaussian Noise)	(None, 99, 160)	0
dropout_1 (Dropout)	(None, 99, 160)	0
LSTM_60_Neurons (Bidirectional LSTM)	(None, 99, 120)	106080
dropout_2 (Dropout)	(None, 99, 120)	0
LSTM_40_Neurons (Bidirectional LSTM)	(None, 99, 80)	51520
dropout_3 (Dropout)	(None, 99, 80)	0
gaussian_noise_2 (Gaussian Noise)	(None, 99, 80)	0
Dense1_Sigmoid (Dense)	(None, 99, 32)	2592
Dense2_Sigmoid (Dense)	(None, 99, 16)	528
flatten (Flatten)	(None, 1584)	0
Output_Softmax (Dense)	(None, 5)	7925
Total params: 228,213		
Trainable params: 228,189		
Non-trainable params: 24		

*Tīkla arhitektūras viettura attēls**Ņemiet vērā, ka divvirzienu LSTM slāņi ir tikai divi paralēli LSTM slāņi, kas divkārtšo slāņa neironu skaitu*

Modeļa arhitektūra ir salīdzinoši maza, atstājot daļu skaitļošanas resursu pieejamus paplašināšanai nākotnē. Tā kā dati ir ļoti specifiski, ir jāsamazina viltus pozitīvo vērtību skaits – šī iemesla dēļ modelis tika koriģēts klasificēt skaņas kā troksni, un 98,47% dzirdamā trokšņa tika identificēts pareizi. Kā norāda neskaidrības matricas (*confusion matrix*) rezultāti, modelim ir grūtības atšķirt klepu no šķaudīšanas. Pārsteidzošāk ir tas, ka testēšanas datu kopā 21,57% šķaudīšanas skaņu tiek klasificētas kā trokšņi, bet tikai 8,50% klepus skaņu tiek klasificētas kā trokšņi. To neapšaubāmi izraisa fakts, ka šķaudīšanas ieraksti satur garas klusuma pauzes vai cita veida trokšņus.

Neskaidrības matricas apmācības un testēšanas datu kopām.



Klasificēto paraugu procentuālā daļa, katrā rindā summējot līdz 100%.

Citi nepareizas klasifikācijas cēloņi, iespējams, ir saistīti ar ievades funkcijas izvēli. *Mel-Frequency Cepstral* koeficients ir īpašības vērtība, kas izgūta no *Mel-Scaled* spektrogrammām, kur *Mel-Scale* ir tuvs cilvēka dzirdes uztveres analogs. *MFCCS* ir viena no populārākajām funkcijām, kas tiek izmantota audioanalīzē un runas atpazīšanā, pat emocionālās noskaņas un patskaņu atpazīšanā. Tā popularitāti veicina arī fakts, ka tas ir ļoti kompakts cilvēka dzirdes spektra attēlojums, samazinot nepieciešamos skaitļošanas resursus un datu kopu uzglabāšanas vietu. *MFCC* tehnoloģija samazina frekvences diapazonu (šajā gadījumā no 0 līdz 1000 Hz) līdz 12 *Cepstral* koeficientiem ar *DCT* kompresijas algoritmu, ko parasti izmanto *JPEG* un *MP3* failu izmēra samazināšanā. Tādējādi tiek zaudēta tāda informācija kā frekvences fāze, un izzūd klepus nianse pret šķaudīšanu.

Ir veikts papildu eksperiments ar sitaminstrumentu skaņu datu kopu. Pievienojot trokšņa kategorijā sitaminstrumentu skaņas, piemēram, dažādas bungas, modelis lielākoties nespēj iemācīties atšķirību starp perkusiju skaņu un šķaudīšanu - līdz 45% šķaņu tiek klasificētas kā troksnis, neatkarīgi no modeļa apmācības ilguma. Tas norāda, ka *MFCC* nevar izmantot šāda veida datiem, un ievades funkcijai nepieciešama augstāka izšķirtspēja. *MFCC* aprēķinu parametri ir šādi: *hop length* vai slidošais logs ar 512 paraugiem, *FFT* izmērs un loga izmērs - 1024 paraugi, 13 *Cepstral* koeficienti, izņemot pirmo jaudas koeficientu un diskretizācijas frekvenci 22050 Hz.

Uzlabotā metode ir padziļināto funkciju (*deep features*) izmantošana, kas iegūtas no *Mel-Scaled* spektrogrammām ar izmēra samazināšanas (*Dimensionality Reduction*) algoritmiem/tīkliem vai *CNN*, kā to iesaka mūsdienu pētījumi [13], [12], [14]. Pēc tam padziļinātās funkcijas var ievadīt rekurentajā neironu tīklā, samazinot kompromisus, kas rodas izmantojot tikai *RNN* vai *CNN*. Lielākos gala-gala (*end-to-end*) neironu tīklos visas augsta līmeņa vai padziļinātās funkcijas tīkls visbiežāk apgūst neatkarīgi. Strādājot ar ierobežotu arhitektūru, ir jēgpilni iekodēt ievades datus pēc apmācīta augstākas dimensijas attēlojuma (*higher-dimensional representation*). Tiek mēģināts ieviest šos uzlabojumus pirms programmatūras galīgās versijas izlaišanas.

Kopsavilkums

Integrēto datoru izmantošanas dēļ bija iespējams izveidot sistēmu ar zemām izmaksām. Masveida uzstādīšanas gadījumā cenu var samazināt vēl vairāk, izmantojot vienu datoru vai mākoņa platformu, lai veiktu aprēķinus no vairākām telpām. Tomēr tam var būt nepieciešami varas iestāžu noteikti papildu regulējumi par privāto datu apstrādi. Bezvadu komunikācija var ievērojami vienkāršot ierīču insolāciju. Izstrādātā sistēma ļauj noteikt cilvēku skaitu, analizēt pamatotu informāciju un izmērīt nepieciešamos vides parametrus, lai iegūtu datus infekcijas riska novērtēšanai. Šādu parametru, piemēram, cilvēku skaita telpā, mērījumus var ievērojami uzlabot, izmantojot papildu pirmapstrādi un tīkla uzlabošanu ar datu apmācību no noteiktas telpas/uzraudzības zonas. Piedāvātais risinājums spēj viegli atšķirt cilvēka balsi un elpošanas skaņas no trokšņa, lai gan atšķirt šķaudīšanas un klepus skaņas ir diezgan sarežģīti. Pieaugot pieprasījumam pēc COVID-19 brīdināšanas sistēmām, iestādes vāc skaņu paraugus, kas ievērojami ļauj uzlabot AI skaņas uzraudzības sistēmas.

Atsauces



- [1] R. Girshick. Fast R-CNN. ICCV, 2015
- [2] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. NIPS, 2015
- [3] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, 2015
- [4] A.Özlu. TensorFlow Object Counting API.
<https://github.com/ahmetozlu/tensorflow-object-counting-api>. 2018
- [5] https://docs.opencv.org/3.4/db/d58/group_calib3d_fisheye.html
- [6] Redmon, Joseph & Farhadi, Ali. (2018). YOLOv3: An Incremental Improvement.
- [7] Li, Yuhong & Zhang, Xiaofan & Chen, Deming. (2018). CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes.
- [8] He, Kaiming & Gkioxari, Georgia & Dollár, Piotr & Girshick, Ross. (2017). Mask R-CNN. 2980-2988. 10.1109/ICCV.2017.322.
- [9] Pang, Jiangmiao & Li, Cong & Shi, Jianping & Xu, Zhihai & Feng, Huajun. (2019). R^2 -CNN: Fast Tiny Object Detection in Large-Scale Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing. 57. 5512-5524. 10.1109/TGRS.2019.2899955.
- [10] Prayitno, Bagas & Suyanto, Suyanto. (2019). Segment Repetition Based on High Amplitude to Enhance a Speech Emotion Recognition. Procedia Computer Science. 157. 420-426. 10.1016/j.procs.2019.08.234.
- [11] Hershey, Shawn & Chaudhuri, Sourish & Ellis, Daniel & Gemmeke, Jort & Jansen, Aren & Moore, R. & Plakal, Manoj & Platt, Devin & Saurous, Rif & Seybold, Bryan & Slaney, Malcolm & Weiss, Ron & Wilson, Kevin. (2017). CNN architectures for large-scale audio classification. 131-135. 10.1109/ICASSP.2017.7952132.
- [12] Y. Li, X. Li, Y. Zhang, M. Liu and W. Wang, "Anomalous Sound Detection Using Deep Audio Representation and a BLSTM Network for Audio Surveillance of Roads," in *IEEE Access*, vol. 6, pp. 58043-58055, 2018, doi: 10.1109/ACCESS.2018.2872931.
- [13] S. Amiriparian, M. Gerczuk, S. Ottl, N. Cummins, S. Pugachevskiy and B. Schuller, "Bag-of-Deep-Features: Noise-Robust Deep Feature Representations for Audio Analysis," *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, 2018, pp. 1-7, doi: 10.1109/IJCNN.2018.8489416.
- [14] Shuvaev, Sergey & Giaffar, Hamza & Koulakov, Alexei. (2017). Representations of Sound in Deep Learning of Audio Features from Music.