TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond

Valeria Juštšenko 203907IABM

Süvaõppel põhinev kõne emotsioonide tuvastamine õppeprotsessis

Magistritöö

Juhendaja: Olga Dunajeva
PhD

Tallinn 2023

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Valeria Juštšenko

03.01.2023

Annotatsioon

Kõnesignaalid on inimeste jaoks üks loomulikumaid suhtlusvahendeid. Närvivõrkude tehnoloogiate kiire arenguga muutub kõneemotsioonide äratundmine üha aktuaalsemaks. Näiteks õppesüsteemis saab tuvastada õpilasi, kellel on igav.

Selle magistritöö eesmärk on luua meetod üliõpilaste ja õppejõudude emotsionaalse seisundi jälgimiseks kõne järgi süvaõppe meetodite abil. Töös uuritakse reaalajas kõnest emotsionaalseisundi määramise võimalust õppeprotsessi käigus. Selleks kasutatakse süvaõppe algoritme ning töö tulemusena luuakse veebirakenduse prototüübi emotsionaalse tausta jälgimiseks, sealhulgas õppeprotsessis. Samuti töös käsitletakse kasutatud andmekogumeid, kirjeldatakse ära andmete eeltöötluse protsess, ja kirjeldatakse parima leitud CNN närvivõrgumudeli arhitektuuri, treenimisprotsessi ja headust. Testandmetel saadud mudeli täpsus on 93.85 %. Lõpus kirjeldatakse saadud mudelil põhinev veebirakenduse prototüüp ja selle arendamise võimalused.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 39 leheküljel, 6 peatükki, 27 joonist, 3 tabelit.

Abstract

Deep Learning Based Speech Emotion Recognition in Educational Process

Speech signals are one of the most natural means of communication for humans. With the rapid development of neural network technologies, the recognition of speech emotions is becoming more and more relevant. For example, in the learning process, students who are bored can be identified.

The purpose of this master's thesis is to create a method for monitoring the emotional state of students and teachers in the learning process using deep learning methods for speech. In the work is considered the possibility of determining the emotional state by speech in real time during the learning process. For this, deep learning algorithms are used, and as a result of the work, a prototype of a web application is created for monitoring the emotional background, including during the learning process. The work also discusses the datasets used, describes the data preprocessing, and describes the architecture of the best CNN neural network model, the training process, and the quality of the resulting model. The accuracy of the model on the test data is 93.85%. At the end, the prototype of the web application based on the obtained model and the possibilities of its development are described.

The thesis is in Estonian and contains 39 pages of text, 6 chapters, 27 figures, 3 tables.

Lühendite ja mõistete sõnastik

CNN	*Convolutional Neural Network*, konvolutsiooniline närvivõrk, mitmekihiliste tehisnärvivõrkude klass
Fourier teisendus	Fourier' teisendus on matemaatiline funktsioon, mida saab kasutada signaali või laine aluseks olevate sageduste leidmiseks.
Hertz (Hz)	Mõõtühik, mis määrab sündmuse korduste arvu sekundis
LSTM	*Long Short Term Memory*, pika lühiajalise mälu närvivõrk, rekurrentse närvivõrgu edasiarendus
MFCC	*Mel-Frequency Cepstral Coefficients,* Mel-skaalal kepstri kordajad, heli sageduslikud tunnusvektorid
MLP	*Multi Layer Perceptron,* mitmekihiline pertseptron on edasisuunaline tehisnärvivõrk, mis genereerib sisendite komplektist väljundite komplekti.
RNN	*Recurrent Neural Network,* rekurrentne närvivõrk
STFT	*Short-Time Fourier Transform,* lühiajaline Fourier' teisendus
Streamlit	Streamlit on avatud lähtekoodiga rakenduste raamistik
Sügavõpe (süvaõpe)	Sügavõpe (*ingl* deep learning) on masinõppe algoritm, mis kasutab inimese ajust inspireeritud mitmekihilisi tehisnärvivõrke, et õppida suurest hulgast andmetest
Tensor	Lineaaralgebra objekt, mis üldistab skalaari, vektori, maatriksi ja bilineaarse vormi mõistet
VAD	*Voice Activity Detection*, helis hääle tuvastuse algoritm
WAV	Meediatöötluses laialdaselt kasutatav formaat tihendamata helivoo salvestamiseks, WAV salvestab mono ja stereo heli erinevate diskreetimissagedustega

Sisukord

Jooniste loetelu

Tabelite loetelu

1 Sissejuhatus

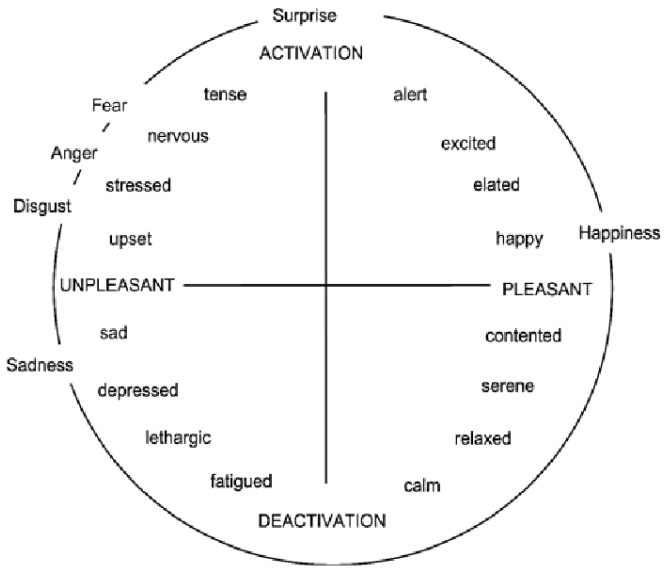
Erinevate uuringute kohaselt on emotsioonide tuvastamine viimastel aastatel palju tähelepanu pälvinud, sellel on oluline roll ka hariduses. Praegu kasutavad õpetajad tagasiside allikana teooriaeksameid ja teste, kuid need meetodid on sageli ebaefektiivsed. Õpilase emotsioonide tuvastamise abil saab õpetaja kohandada oma hindamismeetodeid ja õppematerjale, et hõlbustada õpilaste õppimist.

Käesolevas magistritöös uuritakse TalTech Virumaa kolledži üliõpilaste ja õppejõudude emotsionaalseisundi määramise võimalust õppeprotsessi käigus. Selleks kasutatakse süvaõppe algoritme ning töö tulemusena luuakse veebirakendus õppeprotsessi emotsionaalse tausta jälgimiseks.

1.1 Probleemi taust

Emotsioonid mõjuvad nii inimese füsioloogilist kui ka psühholoogilist seisundit. Positiivsed emotsioonid aitavad parandada inimese tervist ja töö efektiivsust, negatiivsed aga võivad tekitada terviseprobleeme.

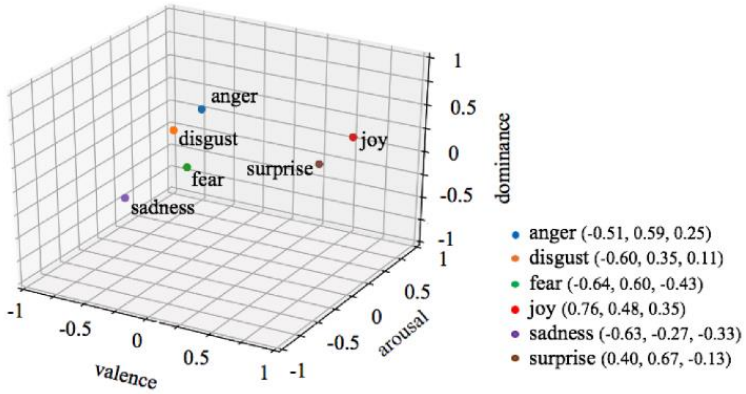
Emotsioonid ja kõne on omavahel tihedalt seotud ning mängivad inimese elus väga olulist rolli. Kuigi emotsioon ise on Russelli teooria [1] kohaselt väga keeruline, nagu näitab Joonis 1, emotsioone saab väljendada kasutades kahte olekut, nagu valents ja aktiveerimine (erutus). Aktiveerimine on füsioloogiline erutus, mis on seotud emotsiooni kogemisega. Aktiveerimine viitab energiale, mis on vajalik selle (kõrge või madala) emotsiooni väljendamiseks. Joonisel valents on esindatud negatiivsest positiivseni, samal ajal kui aktiveerimine on esindatud erutuse puudumisest erutuseni või nõrgast intensiivseni. Tugevamad emotsioonid on viha, rõõm, hirm. Selliste tundmustega võib kaasneda südamepekslemine, vererõhu tõus jne. Samal ajal suureneb inimeste kõne kiirus, samuti muutub helikõrgus kõrgemaks. Rahustavamad emotsioonid, nagu kurbus, vastupidi, vähendavad kõne kiirust ja sagedust. Aga sarnaste tunnete aktiveerimine, nagu viha ja rõõm, erineb valentsi poolest. Valents tähendab positiivseid või negatiivseid emotsioone [2].



Joonis 1. Russelli kahemõõtmeline emotsioonide mudel[1]

Emotsioonide klassifitseerimiseks on olemas ka mitmemõõtmelised mudelid, eeldades, et iga üksikemotsioon on mitme mõõtme kombinatsioon. Näiteks 3D-mudel, mis on näidatud pildil (Joonis 2) käsitleb valentsi, erutust ja domineerimist.

Mõiste "domineerimine" näitab emotsiooni domineerivat olemust ja määrab inimesele mõjutava astme. Siin võib näiteks tuua viha ja hirmu, mis ei paku naudingut, kuid asuvad domineerimise mõttes vastaskülgedel.



Joonis 2. Emotsioonide 3D-mudel. [2]

[1] https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7319769&tag=1

[2] https://www.researchgate.net/publication/352676025_Emotion_Recognition_From_EEG_Signal_Focusing_on_Deep_Learning_and_Shallow_Learning_Techniques

Seega eristuvad tugevad emotsioonid hästi rahustavatest emotsioonidest, kuid eri tüüpi emotsioone on siiski raske eristada.

Mainida tuleb ka peamisi takistusi emotsioonide tuvastamisel kõnes (SER - Speech Emotion Recognition).

1. Isegi inimesed interpreteerivad emotsioone erinevalt. Ise "emotsiooni" mõistet on raske määratleda.

2. Andmestikute loomisel on heli märkuste tegemine keeruline. Jäävad küsimused: kas on vaja kuidagi tähistada iga üksikut sõna, lauset või kogu suhtlust tervikuna? Millist emotsioonide kogumit tuleks äratundmiseks kasutada?

3. Ka andmete kogumine pole lihtne. Filmidest ja uudistest saab koguda palju heliandmeid. Mõlemad allikad on aga "ebaobjektiivsed", sest uudised peavad olema neutraalsed, näitlejate emotsioonid on mängitud. Heliandmete "objektiivset" allikat on raske leida.

4. Andmete märgendamine nõuab suuri inim- ja ajaressursse. Tervete helisalvestiste kuulamiseks, analüüsimiseks ja kommentaaride esitamiseks on vaja spetsiaalselt koolitatud personali. Ja siis peavad paljud teised inimesed neid kommentaare hindama, sest hinnangud on subjektiivsed. [3]

1.2 Eesmärk ja ülesanded

Selle magistritöö eesmärk on luua meetod õppeprotsessi käigus üliõpilaste ja õppejõudude emotsionaalse seisundi jälgimiseks kõne järgi süvaõppe meetodite abil.

Töös püstitatud eesmärgi saavutamiseks on vaja lahendada järgmised ülesanded:

- Uurida teoreetilist materjali süvanärvivõrkude treenimise ja nende omaduste kohta seoses kõnetöötlusega.
- Viia läbi viimase viie aasta jooksul avaldatud publikatsioonide analüüs, mis on seotud kõne emotsioonide tuvastamisega õppeprotsessis.
- Leida emotsioonide tuvastamiseks võimalikult efektiivne mudel üliõpilaste ja õppejõudude emotsionaalse seisundi määramiseks õppeprotsessi käigus.

- Luua veebirakenduse prototüüp üliõpilaste ja õppejõudude emotsionaalse seisundi jälgimiseks õppeprotsessis.

1.3 Ülevaade tööst

Töö teises peatükis antakse ülevaade teemaga seonduvast kirjandusest. Kolmandas jaotises käsitletakse kasutatud andmekogumeid ja hoidlaid. Töö neljas osa keskendub töö käigus kasutatud vahendite ja tööriistade kirjeldamisele, kirjeldatakse ära andmete eeltöötluse protsess, käsitletud mudelid ja nende valideerimise tehnikad. Jaotises 5 kirjeldatakse närvivõrgumudeli arhitektuuri, treenimisprotsessi ja saadud mudeli headust. Samuti esitatakse saadud mudelil põhinev veebirakenduse prototüüp ning analüüsitakse saadud tulemusi ja käsitletakse ka võimalusi edasiseks tööks.

2 Varasema kirjanduse ülevaade

Hiljutised edusammud närvivõrkude alas on viinud nende edukate rakendusteni peaaegu kõigis inimelu aspektides. Emotsioonidel on igapäevastes inimestevahelises suhtluses oluline roll. See aitab teiste tundeid mõista, edastades oma tundeid ja andes teistele tagasisidet. Emotsionaalsed esinemisvormid annavad olulist teavet inimese vaimse seisundi kohta. See on avanud uue uurimisvaldkonna, mida nimetatakse automaatseks emotsioonituvastuseks, mille peamine eesmärk on mõista ja tekitada soovitud emotsioone.

See peatükk annab ülevaate olemasolevatest uuringutest, mis käsitlevad emotsioonide tuvastamist kõnes, kasutades erinevaid tehnikaid. Nende uurimistööde eesmärgid on leida parimad meetodid emotsioonide tuvastamiseks.

2022. aastal ilmunud artiklis [4] Husbaan I. Attar ja Nilesh K. uurisid pideva kõne emotsioonide tuvastamist ja pakkusid välja ka reaalajas kõneemotsioonide tuvastamise süsteemi veebipõhiseks õppeks. Tulemus väljendub neljas erinevas emotsioonikategoorias: viha, kurbus, õnn, neutraalne. Süsteem koosneb hääletegevuse tuvastamisest, kõne segmenteerimisest, signaali eeltöötlusest, tunnuste eraldamisest, emotsioonide klassifikatsioonist ja emotsioonide sageduse statistilisest analüüsist. Analüüs viidi läbi RAVDESS andmebaasi helisalvestistel. Konvolutsioonilise närvivõrgu (CNN) mudeli loomiseks ekstraheeriti Mel-cepstrali koefitsiendid (MFCC). Katsed viidi läbi nii eelsalvestatud andmekogumitega kui ka reaalajas salvestamisega. Katsete keskmine täpsus on 90% ja 78,78% vastavalt. Simuleeritud veebipõhises õpikeskkonnas läbiviidud katse tulemused näitavad, et süsteem suudab tõhusalt tuvastada õpilase reaktsiooni kursusele ja õppimisele ning aidata õpilastel saavutada optimaalset õppeedukust.

Ling Cen jt [5] pakkusid välja online-õppel põhineva emotsioonide tuvastamise süsteemi. Süsteemi põhifunktsioonid: kõne algus- ja lõpp-punktide tuvastamine, reaalajas kõne salvestamine läbi mikrofoni, helifaili loomine ja laadimine, töödeldava kõnesignaali lainekuju kuvamine ja pideva kõne automaatne segmenteerimine, emotsionaalsete seisundite tuvastamine helifailist või pidevast kõnest reaalajas, tuvastamise tulemuste kuvamine, emotsioonide sageduse statistiline analüüs kõneandmete kogumi jaoks või reaalajas. Kõnesignaali eeltöötluseks kasutati kadreerimist ja aknaid, millele järgneb

tunnuste eraldamine. Kepstri kordajate tunnused eraldatakse töödeldud signaalist: lineaarsel ennustamisel põhinevad kepstri kordajad (Linear Predictive Cepstral Coefficients, LPCC); tajutud lineaarse ennustuse kepstri kordajad (Perceptual Linear Prediction, PLP); ja mel-sageduse kepstri kordajad (Mel-frequency Cepstral Coefficients, MFCC). Mudel on üles ehitatud tunnuste ja emotsioonikategooriate sihtväärtuste sisestamisel tugivektormasina (SVM) algoritmi. Emotsioonide tuvastamise täpsus saadud mudeli kasutamisel jaotub erinevalt keskmise väärtusega 90%.

Anusha jt [6] käsitlesid RAVDESSi andmestikku. Mudeli treenimiseks eraldati kolm põhifunktsiooni, nagu MFCC (MelFrequency Cepstral Coefficients), Mel spektrogramm ja kromaatilisus. Klassifitseerimiseks süsteemis kasutatati MLP klassifikaatorit, mille täpsus oli umbes 80%.

Chen Jin jt [7] eraldasid helifailidest Mel-sageduse kepstri kordajad ning ehitasid ühemõõtmelise sügava jääknärvivõrgu mudeli, mis põhineb konvolutsioonilisel närvivõrgul. Mudeli tuvastustäpsus valideerimisandmetel Savee ja RAVDESS-i andmestikutel on 98,26% ja 97,30% vastavalt. Mudel kasutab RMSPropi optimeerijat, õppimiskiiruseks on 0,0001 ja sumbumiskiiruseks 1e-6, kaofunktsiooniks on categorical_crossentropy.

L. Jie jt [8] uurisid õpetajate kõnesignaale ja töötasid välja helitöötlussüsteemide komplekti emotsioonide tuvastamiseks. Emotsioonide hindamiseks kasutati õpetajate kõnet. Kõneemotsioonide tuvastamise klassifikatsioonimudeli koostamiseks kasutati rekurrentse närvivõrgu (RNN) algoritmi. Tunnuste eraldamisprotsessis kasutati 39- ja 40-meerilisi Mel Frequency Cepstral Coefficents (MFCC). Selle tulemusena on viie kõneemotsiooni keskmine tuvastamistäpsus 39-mõõtmelise tunnuste parameetrite puhul 75,64% ja 40-mõõtmelise puhul - 78,17%. Katsete tulemusena saadi ka pikaajalise lühiajalise mälu optimeerimise mudel (LSTM) viie tüüpi kõneemotsioonide tuvastamistäpsusega 85,32%. Emotsioonide klassifitseerimiseks kogutakse juhendaja kõnesignaali ning luuakse õpetaja kõneemotsioonide andmebaas, mis võimaldab automaatselt hinnata reaalajas õppetöö kvaliteeti klassiruumis.

Ristea jt [9] pakkusid välja kasutada sügavatel konvolutsioonilistel närvivõrkudel põhinevat süsteemi, mis suudab emotsioone reaalajas tuvastada. Tuvastamissüsteemi täpsuse parandamiseks kombineeritakse visuaalsetest ja heliallikatest pärinev teave.

Andmestiku CREMA-D katsetulemused näitavad visuaalsete andmete heliga kombineerimise tähtsust. Ainult visuaalsete andmete kasutamise korral on täpsus 62,84% ning kombineeritud andmete: heli ja video kasutamisel on täpsus 69,42%.

3 Andmestikud

Andmestikud mängivad automaatses emotsioonide tuvastamises olulist rolli, kuna mudelid õpivad näidetest. Mudelite treenimiseks kasutatavad andmebaasid võivad sisaldada nii spetsiaalselt koostatud kui ka reaalseid emotsioonide avaldamise näiteid. Andmebaasi loomulikkuse kasvades suureneb ka keerukus. Nii tehti vokaalsete emotsioonide automaatse äratundmise uuringute alguses, need uuringud algasid 90-ndate keskel. Baasi täitmine algas näitlejakõnega ja on nüüdseks nihkumas realistlikumate andmete poole [10].

Praegu pole emotsioone kirjeldavate häälsõnumitega palju andmeid avalikus omandis.

Mudeli treenimiseks leidis autor 5 andmekogumit, mida saab uuringus kasutada: TESS, SAVEE, RAVDESS, CREMA ja EMO.

3.1 TESS andmestik

TESS andmestik[1] sisaldab 2800 WAV-heliriba. 200 sihtsõnast koosnevat komplekti räägiti kandefraasis "Say the word _____". Seda andmestikku helindavad kaks naishäält ja sellele on lisatud 7 emotsiooni: viha, vastikus, hirm, õnn, kurbus, üllatus, neutraalne emotsioon. Failinimi näitab, millist sõna öeldakse ja millise emotsiooniga. Näiteks "OAF_back_fear.wav" on esimene hääl, mis ütleb fraasi "Say the word back" koos emotsiooniga "hirm".

3.2 SAVEE andmestik

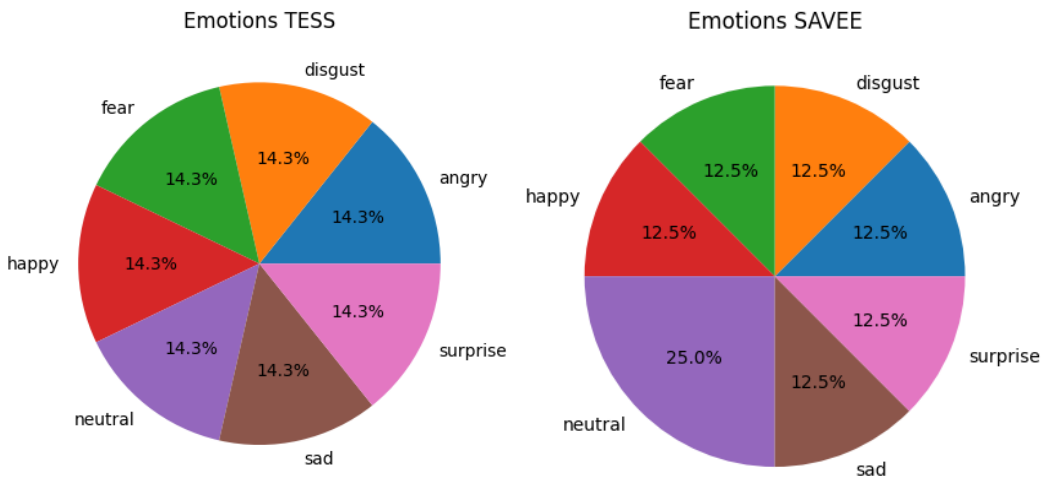
SAVEE andmestikku[2] helindavad neli meesnäitlejat ja see on märgistatud 7 erineva emotsiooniga. Andmestik sisaldab 480 WAV-heliriba. Seal on 15 lauset iga 7 emotsioonikategooria kohta.

[1] https://borealisdata.ca/dataset.xhtml?persistentId=doi:10.5683/SP2/E8H2MF

[2] http://kahlan.eps.surrey.ac.uk/savee/Database.html

Andmestikus on WAV-failid diskreetimissagedusega 44,1 kHz. Failinime algustäht(ed) tähendavad emotsiooniklassi ja järgmised numbrid tähendavad lause numbrit. Tähed 'a', 'd', 'f', 'h', 'n', 'sa' ja 'su' tähendavad emotsiooniklasse "viha", "vastikus", "hirm", "õnn", "neutraalne", "kurbus" ja "üllatus". Näiteks "d03.wav" on kolmas lause, millel on "vastikus" emotsioon.

Emotsionaalsete seisundite jaotus TESS ja SAVEE andmekogumites näidatud joonisel (Joonis 3).



Joonis 3. Emotsionaalsete seisundite jaotus TESS ja SAVEE andmekogumites.

3.3 RAVDESS andmestik

Andmebaas RAVDESS[1] (The Ryerson Audio-Visual Database of Emotional Speech and Song) sisaldab 24 professionaalse näitleja salvestusi. Neist 12 naist ja 12 meest esitasid kaks väidet (1 = "Kids are talking by the door", 2 = "Dogs are sitting by the door") neutraalse Põhja-Ameerika aktsendiga. Kõne sisaldab emotsioone: rahulik, õnn, kurbus, viha, hirm, üllatus ja vastikus ning lisaks neutraalse emotsiooniga. Iga väljend on loodud kahel emotsionaalse intensiivsuse tasemel (normaalne, tugev). Heli on saadaval wav-vormingus (16bit, 48kHz). Andmestik sisaldab 1440 faili: 60 salvestust näitleja kohta.

Igal failil on kordumatu nimi. Faili nimi koosneb 7-osalisest numbrilisest identifikaatorist. Emotsioonide märgistused on 01 = neutraalne, 02 = rahulik, 03 = õnn, 04 = kurbus, 05 = viha, 06 = hirm, 07 = vastikus, 08 = üllatus. Faili nime näide: 03-01-

[1] https://zenodo.org/record/1188976

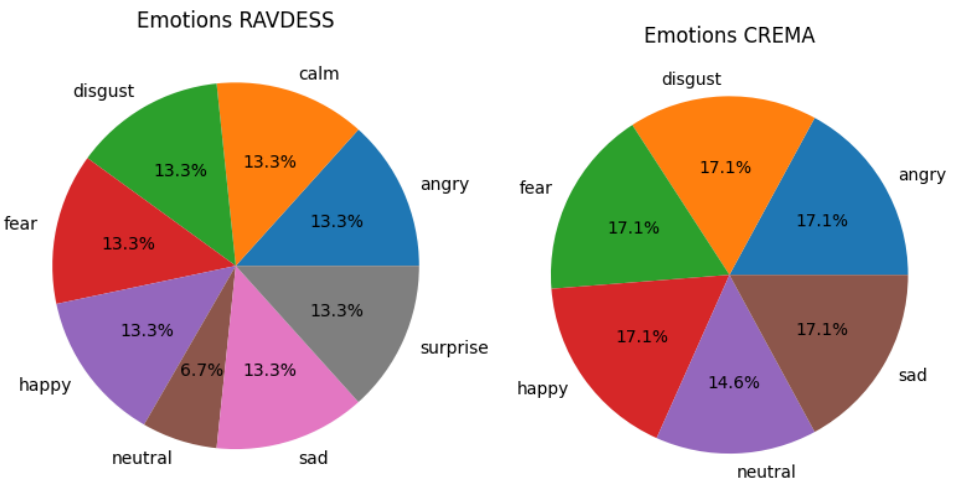
06-01-02-01-12.wav - ainult heli (03), kõne (01), hirm (06), normaalne intensiivsus (01), väljend " Dogs are sitting by the door" (02), 1. kordus (01), 12. näitleja (12) - naishääl, kuna näitleja ID number on paaris (paaritu numbriga näitlejad on mehed).

3.4 CREMA andmestik

CREMA-D[1] andmebaas ilmus 2015. aastal ja sisaldab 7442 klippi 91 erineva etnilise päritoluga näitlejast (48 meest ja 43 naist), mille on filminud professionaalsed teatrijuhid. Näitlejatel paluti edastada konkreetseid emotsioone, öeldes 12 konkreetset lauset erinevates intonatsioonides, mis tekitasid sihtemotsiooni. Seal on kuus märgistatud emotsiooni ja neli erinevat emotsiooni taset.

Näitleja ID on 4-kohaline number faili alguses. Iga järgmine identifikaator on eraldatud alakriipsuga (_). Failinime teises osas kasutatud kolmetäheline lühend määrab kasutatud lauset. Laused esitati 6 erinevat emotsiooni kasutades (failinime kolmandas osas kolmetäheline kood): viha (ANG), vastikus (DIS), hirm (FEA), õnn (HAP), neutraalne (NEU), kurbus (SAD) ja 4 emotsioonitaset (kahetäheline kood failinime neljandas osas): madal (LO), keskmine (MD), kõrge (HI), täpsustamata (XX).

Emotsionaalsete seisundite jaotus RAVDESS ja CREMA andmekogumites on näidatud joonisel (Joonis 4).



Joonis 4. Emotsionaalsete seisundite jaotus RAVDESS ja CREMA andmekogumites.

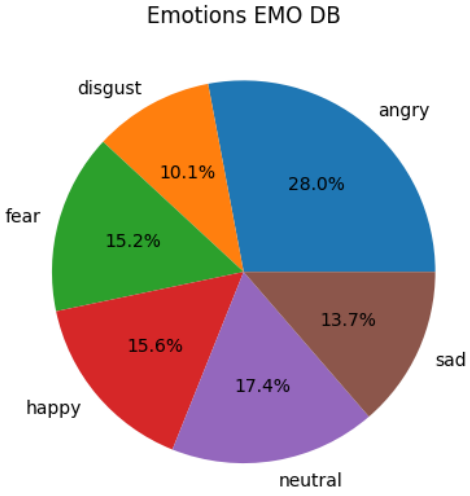
[1] https://github.com/CheyneyComputerScience/CREMA-D

3.5 EMO andmestik

Berliini emotsionaalse kõne andmebaas[1] sisaldab umbes 500 näitlejate helisalvestusi rõõmsates, vihastes, ärevates, hirmunud, igavates, vastikutes ja neutraalsetes versioonides. Väited salvestasid 10 erinevat näitlejat kasutades 10 erinevat teksti.

Failinimed on sama skeemi järgi: positsioonid 1-2: kõneleja number, positsioonid 3-5: teksti kood, positsioon 6: emotsioon (saksa täht), positsioon 7: kui versioone on rohkem kui kaks, on need nummerdatud a, b, c. Emotsioonikoodid (saksa täht): W = viha, L = igavus, E = vastikus, A = hirm, F= õnn T = kurbus N = neutraalne versioon. Näide: 03a01Fa.wav on helifail, milles kolmas kõneleja ütleb teksti a01 emotsiooniga "Freude" (Õnn). Teave esinejate kohta - 08, 09, 13, 14, 16 - naine, 03, 10, 11, 12, 15 - mees.

Emotsionaalsete seisundite jaotus EMI andmekogumis näidatud joonisel (Joonis 5).



Joonis 5 EMO andmestiku emotsioonide jaotus.

3.6 Andmestikute ühendamine

Väljatöötatud süsteemi efektiivsuse tõstmiseks ühendas autor kõik andmestikud üheks.

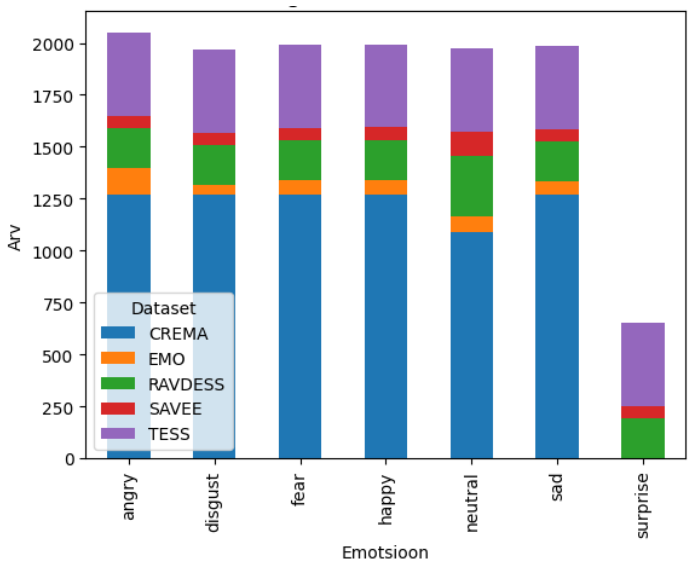
Kuna RAVDESS andmestikus „Calm“ (rahu) emotsiooni on väga vähe andmeid (ainult 921), otsustas autor selle ühendada "neutraalse" emotsiooniga.

[1] http://emodb.bilderbar.info/docu/

Seega on lõplik andmestik, mis koosneb 12616 wav-heliribast, kus on nii mees- kui ka naishääled ning markeeritud 7 emotsiooniga: neutraalne emotsioon, õnn, kurbus, viha, hirm, vastikus, üllatus. Tabel 1 ja Joonis 6 esitavad klasside jaotuse lõplikus andmestikus.

Tabel 1. Emotsioonide järgi jaotus

Emotion	Speech Sample Count
angry	2050
happy	1994
fear	1992
sad	1985
disgust	1969
neutral	1974
surprise	652
Total	**12616**



Joonis 6. Emotsioonide jaotus lõplikus andmestikus.

4 Metoodika ja kasutatud tehnoloogiad

4.1 Kasutatud tarkvara ja riistvara

Arvuti riistvara:

- Protsessor: Intel® Core™ i7-11700 (8 tuuma, 16 lõime) 2.50GHz
- Graafikakaart: NVIDIA GeForce GTX 1660
- Muutmälu: 32 GiB

Kasutatud tarkvara:

Töö teostamiseks oli valitud programmeerimiskeel Python, kuna sellel keelel on suur hulk väliste teeke närvivõrkude arendamiseks ja andmeanalüüsiks. See on kõige arenenum masinõppekeel, kusjuures pakub töötamiseks lihtsat süntaksit.

Koodi kirjutamiseks oli kasutatud VS Code arenduskeskkonda.

Librosa on Pythoni programmeerimiskeele teek, mis on loodud helisalvestiste töötlemiseks ja analüüsimiseks. Võimaldab töödelda heli aegridade kujul, eraldada tunnuseid: tempo, löök, takt, intervall, rütm ja töötada Mel-cepstrali koefitsientidega (MFCC) [11].

NumPy on Pythoni programmeerimiskeele avatud lähtekoodiga teek, mis on loodud töötamiseks mitmemõõtmeliste massiividega ja kõrgetasemeliste matemaatiliste funktsioonidega. See teek kasutatakse helifailide eeltöötluseks, samuti andmete laadimiseks ja töötlemiseks. [12]

Pandas on Pythoni programmeerimisteek andmete töötlemiseks ja analüüsimiseks. Tema abiga viiakse läbi helifailide andmete töötlemine ja andmestiku teabe analüüs.

Matplotlib – võimaldab koostada erinevaid diagramme, sh. Spektrogrammid.

Keras on avatud lähtekoodiga teek, mida kasutatakse süva- ja masinõppe jaoks [13].

Scikit-Learn (sklearn) on üks vanimaid ja suhteliselt üldotstarbelisi Pythoni masinõppe teeke. See sisaldab suure hulga lihtsasti kasutatavaid (valmiskujul) vahendeid ja algoritme, mis on koondatud vastavatesse klassidesse ja moodulitesse [14].

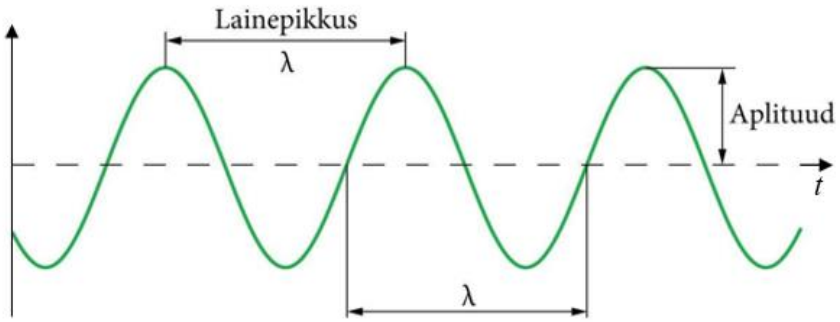
4.2 Andmete eeltöötlus

Inimese jaoks on kõne üks loomulikumaid eneseväljendusviise. Kõne on helide jada. Heli on omakorda erinevate sageduste helivibratsioonide (lainete) superpositsioon. Füüsikast teadaolevalt iseloomustab lainet kaks atribuuti - amplituud ja võnkesagedus (Joonis 7). Mida suurem on heli amplituud, seda valjem see heli inimese jaoks on, mida suurem on sagedus, seda kõrgem on toon.

Amplituud on võnkekõvera harja kaugus keskjoonest ja iseloomustab helitugevust [15].

Lainepikkus on kahe järjestikuse tsükli analoogiliste punktide vaheline kaugus [15].

Tsükli ajalist kestust nimetatakse perioodiks [16].



Joonis 7. Lainepikkus ja amplituud. [1]

Sagedus on füüsikaline suurus, perioodilise protsessi tunnus, mis on võrdsete ajavahemike tagant korduvate sündmuste arv ajaühikus. Sageduse ühikuks on herts (Hz) [17].

Kõik looduses olevad signaalid on sisuliselt analoogsed. Digitaalse signaali töötlemiseks, salvestamiseks ja edastamiseks digitaalsel kujul on analoogsignaalid eelnevalt digitaliseeritakse.

[1] https://opik.fyysika.ee/index.php/book/section/3693#/section/3693

Heli digitaliseerimine on tehnoloogia, mille abil muundatakse analoogne helisignaal digitaalseks, mõõtes signaali amplituudi konkreetsel ajasammul ja salvestades seejärel väärtused numbrilisel kujul [18].

Heli digiteerimine hõlmab kahte protsessi: diskreetimise protsessi ja amplituudi kvantimise protsessi.

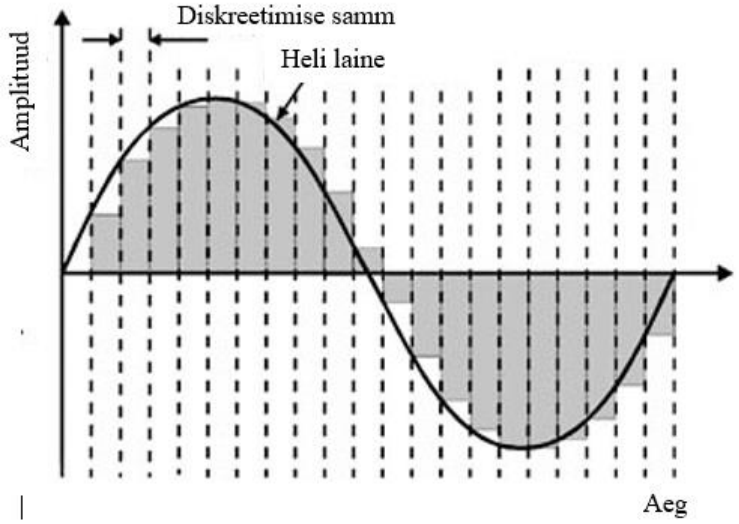
Diskreetimise protsess on analoogsignaali amplituudi väärtuste saamise protsess teatud kindla ajasammuga – diskreetimise sammu [18].

Diskreetimissagedus (*ingl*. *sample rate*) määrab signaali diskreetimisel saadud üksikväärtuste arvu sekundis ja mõõdetakse hertsides (Hz) [19].

Mida väiksem on diskreetimise samm, seda suurem on diskreetimissagedus ja seda täpsem on signaali esitus [20].

Pideva analoogsignaali diskreetimine valitud diskreetimissagedusega on näidatud joonisel (Joonis 8).

Amplituudi kvantimine on protsess, mille käigus asendatakse signaali amplituudi tegelikud väärtused teatud täpsusega ligikaudsete väärtustega [21].



Joonis 8. Heli diskreetimissagedus. [1]

[1] https://3dnews.ru/170037/page-1.html

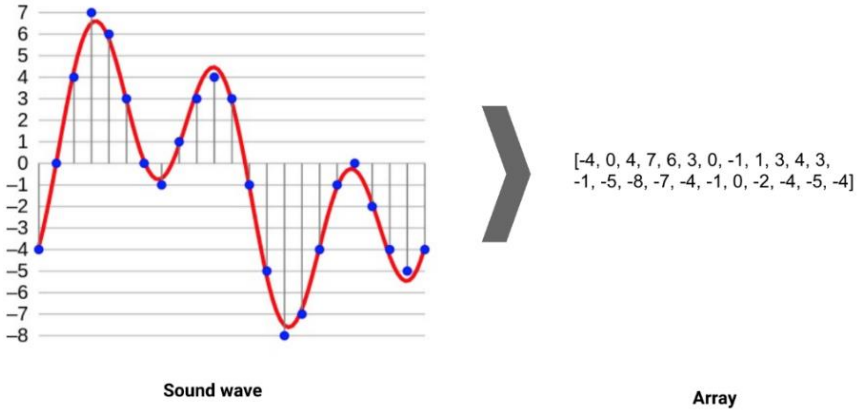
4.2.1 Andmete analüüs

Erinevate emotsioonide helisignaalide omaduste väljaselgitamiseks saab koostada jooniagramme ja spektrogramme.

Helisignaali aegread võimaldavad visualiseerida heli tugevust antud ajahetkel.

Librosa pakett oli kasutatud helifailide andmete analüüsimiseks ja eraldamiseks.

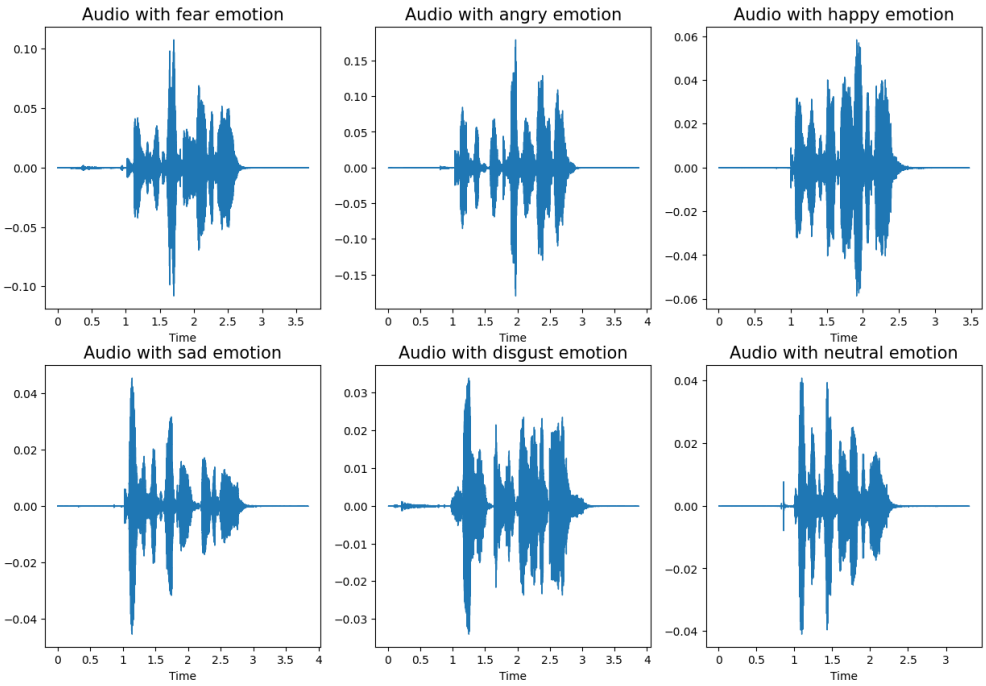
Heli aegridade väärtuste saamine massiivina (Joonis 9) koos diskreetimissagedusega on tehtud funktsiooni load() abil: y , sample_rate = librosa.load(file).



Joonis 9. Helisignaali aegrea väärtused loendis [26].

Saadud massiivi visualiseerimiseks kasutatakse matplotlib teegi alammoodulit pyplot ja *librosa.display.waveplot()* funktsiooni. Graafikute joonistamise kood on esitatud Lisas 1.

Näiteks hirmu, viha ja õnne, kurbuse, vastikuse ja neutraalse emotsiooni lained on näidatud joonisel (Joonis 10).



Joonis 10. Erinevate emotsioonide lained.

Ilmselgelt jõuavad viha, hirmu ja õnne emotsioonid suurema amplituudini 0,5, sest inimesed väljendavad selliseid emotsioone tavaliselt valjult. Kuid see näitab ka, et need emotsioonid kattuvad ajas märkimisväärselt. Sellised emotsioonid nagu kurbus, neutraalsus, vastikus on madalama amplituudiga (< 0,4). Nendes emotsioonides on ka rohkem hingamismüra võrreldes hirmu ja vihaga. Seetõttu on kõne ja emotsioonide tuvastamiseks sobivamad spektraalsed tunnused, kuna need on ajavaldkonnas selliste tunnuste suhtes muutumatud.

Spektrogrammid kujutavad endast aja jooksul muutuva helisignaali sagedusspektri visuaalset esitust [22].

Spektrogrammi levinuim esitus on kahemõõtmeline diagramm: horisontaaltelg tähistab aega, vertikaaltelg - sagedust; kolmas mõõde näitab amplituudi kindlal sagedusel kindlal ajahetkel ja seda esindab pildi iga punkti intensiivsus või värv [22].

Fourier teisendusi (*ingl. Fourier Transforms*) kasutatakse ajapiirkonna teisendamiseks sageduspiirkonnaks. Aja piirkond näitab, kuidas signaal aja jooksul muutub. Sageduspiirkond näitab, kui suur osa signaalist on sagedusvahemikus igas antud sagedusalas [23]. Sagedusala näitab, kui suur osa signaalist on igas määratletud sagedusribas sagedusvahemikus.

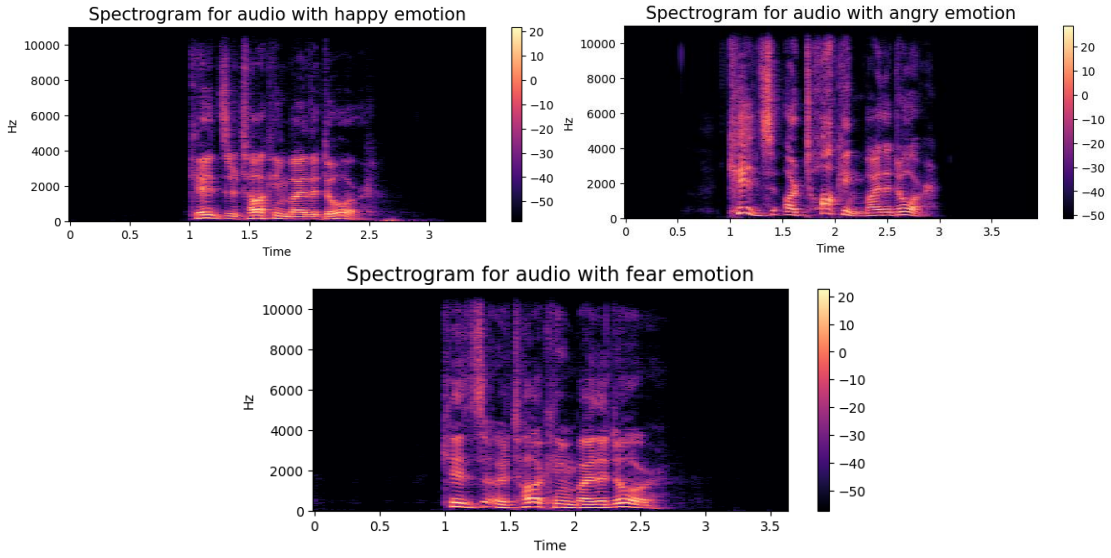
Kõige levinum meetod spektrogrammi arvutamiseks on diskreetne lühiajaline Fourier' teisendus (STFT), mida kirjeldatakse järgmise valemiga (1):

$$STFT\{x[n]\}(m,k) = \sum_{m=-\infty}^{\infty} x[m] \cdot w[n-m] e^{-j\frac{2\pi}{N_x}kn} \tag{1}$$

kus N_x on mõõtmiste arv [31].

Spektrogrammi koostamiseks kasutatakse librosa.display.specshow() (lingid koodile on Lisas 2).

Joonis 11 näitab "õnne", "viha" ja "hirmu" emotsioonide spektrogramme.



Joonis 11. Spektrogrammid.

4.2.2 Tunnuste eraldamine

Helifailide eeltöötluse põhiülesanne on nendest tunnuste eraldamine. Selleks tasub kaaluda mõningaid kontseptsioone, et valida vajalikud funktsioonid [23].

Uuringud on näidanud, et inimese kõrv on madalatel sagedustel helimuutuste suhtes tundlikum kui kõrgetel sagedustel. Sellega seoses võeti kasutusele uus helikõrguse mõõtühik - mel. See põhineb inimese psühhofüsioloogilisel helitajul ja sõltub logaritmiliselt sagedusest. Seega kasutatakse kõneanalüüsiks praktikas tavaliselt mel-spektrogramme [22].

Mel-spektrogramm (MEL Spectrogram Frequency) on tavaline spektrogramm, kus sagedust väljendatakse mitte Hz, vaid Mel kaudu. Hertside teisendamine mel-iks toimub valemi (2) järgi [22]:

$$mel = 1127 \cdot ln\left(1 + \frac{freq}{700}\right) \quad (2)$$

Librosa teegi abil saab koostada mel-spektrogrammi: librosa.feature.melspectrogram(y).

Mel-skaalal kepstri kordajad (*Mel-Frequency Cepstral Coefficients,* MFCC) on helitöötluse üks olulisemaid omadusi. MFCC iseloomustavad inimhääle tämbrilisi aspekte. Need on väike kogum funktsioone (tavaliselt umbes 10–20) [24-25].

MFCC-d saab hõlpsasti ekstraheerida librosa teegi abil: librosa.feature.mfcc(y).

Värvsus või kromaatilisust (*ingl. Chroma feature*) on tavaliselt 12 elemendist koosnev tunnusvektor, mis näitab, kui palju energiat on iga helikõrguse klassist helisignaalis olemas. Arvutatakse logaritmitud sageduse amplituudi spektri summeerimisel oktaavide kaupa. Librosa-s see leitakse funktsiooniga librosa.feature.chroma_stft(y) ning visualiseeritakse kromagrammil [26].

Kontrastsus (*ingl. Spectral Contrast*) librosa-s arvutatakse funktsiooniga librosa.feature.spectral_contrast(). Kontrastsuse kõrged väärtused vastavad üldiselt selgetele kitsaribalistele signaalidele, samas kui madalad kontrastsuse väärtused vastavad lairibamürale [26].

Tonettid ehk helitoonide võrgustik (*ingl. tonnetz*) - tonaalsed tsentroidid ehk "kesksed" toonid - edastavad helisignaali harmoonilist sisu ning aitavad tuvastada heli harmoonilisi muutusi või heli toonidest hälbeid. Librosa-s arvutatakse funktsiooniga librosa.feature.tonnetz(y) [26].

Nulli ületamise määr (*ingl. Zero-Crossing Rate, ZCR*) näitab, mitu korda signaal ületab horisontaaltelje, st mitu korda jõuab amplituud 0-punktini. Nulli ületamise määr on lihtne viis signaali sujuvuse mõõtmiseks. Seda saab arvutada librosa funktsiooniga librosa.feature.zero_crossing_rate(y) [26].

Ruutkeskmine energia (*Root Mean Square Energy, RMS*) – iga kaadri ruutkeskmine väärtus [27].

Signaali koguamplituud vastab signaali energiale. Helisignaalide puhul vastab see tavaliselt signaali valjusele. Signaali energia arvutatakse järgmiselt [27]:

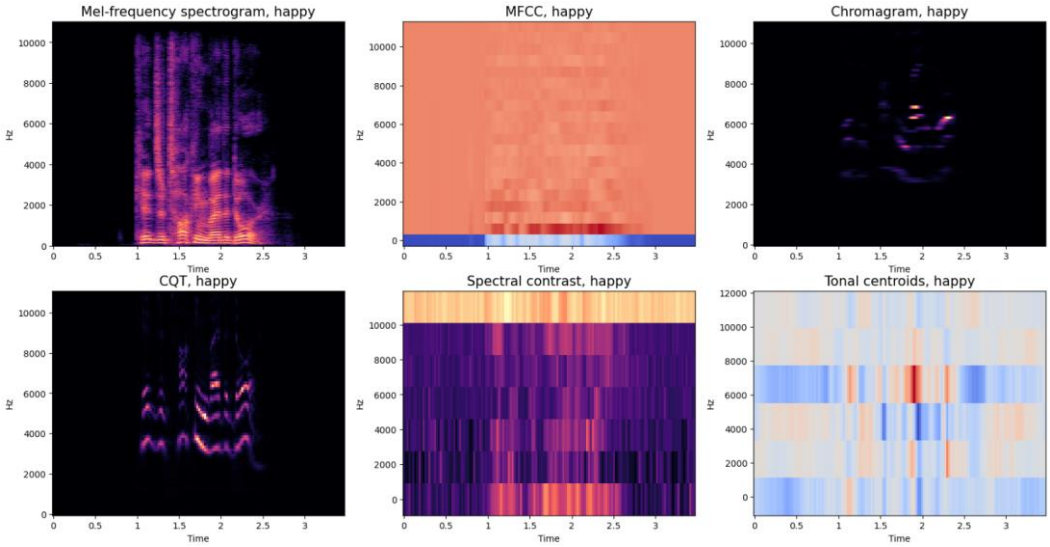
$$\sum_{n} |x(n)|^2 \tag{3}$$

RMSE — kasulik meetod muutujate keskmise väärtuse arvutamiseks ajas. Heliga töötades ruudustatakse signaali väärtus (amplituud), keskmistatakse aja järgi ja seejärel leitakse tulemuse ruutjuur. Signaalienergia ruutkeskmise (RMSE) arvutamise valem [28]:

$$RMSE(x) = \sqrt{\frac{1}{N}\sum_{n} |x(n)|^2}\ \sqrt{\frac{1}{N}\sum_{n} |x(n)|^2} \tag{4}$$

Seda saab arvutada librosa teegi abil funktsiooniga librosa.feature.rms(x).

Joonisel 12 on esitatud audiofailist eraldatud tunnuste diagrammid emotsiooni “õnne” korral.



Joonis 12. Tunnuste diagrammid.

4.2.3 Andmete rikastamine

Andmete rikastamine (*ingl. data augmentation*) on laialdaselt kasutatav strateegia masinõppe mudelite treenimiseks. See teeb osaliselt lihtsamaks piiratud andmete probleemi selliste ülesannete jaoks nagu kõne emotsioonide tuvastamine (SER), kus andmete kogumine on kulukas ja keeruline [28].

Andmete rikastamise all mõistetakse õppimiseks kasutatava andmevalimi suurendamist olemasolevate andmete muutmise kaudu [29].

Andmete rikastamisel on mitu eesmärki nagu treeningvalimi ja kogu andmebaasi suurendamine, mudelite täpsuse ja mürakindluse parandamine [30].

Andmete rikastamiseks saab paremate treeningtulemuste saavutamiseks kasutada selliseid funktsioone nagu müra tekitamine või helisignaalide ajas venitamine. Heliandmete genereerimiseks saab rakendada ajanihet, helikõrguse ja kiiruse muutmist [31].

Müra tekitamine (*ingl*. *noise*) - andmetele juhuslike väärtuste lisamine.

Helikõrguse nihutamine (*ingl*. *pitch shifting*) - kasutatakse helikõrguse tõstmiseks või langetamiseks helisignaali ajas venitamise ja *resampling*u (helifaili diskreetimissageduse muutmine) kaudu. [32]

Venitus (*ingl*. *stretching*) - venitab aegrida (teeb ajalist tihendamist/laiendamist) fikseeritud suurusega jättes signaali helikõrguse muutmata, kuid muudab selle kiirust (tempot). See on kasulik, kui on vaja muuta hääle kiirust ilma hääle tämbrit muutmata. [33]

Ajanihe (*ingl*. *time shifting*) - aegrea nihutamine mõne suvalise sammu võrra tahapoole või ettepoole [34].

Valimi rikastamine tehakse enne Mel-skaala kepstri kordajate arvutamist. Parima efekti saavutamiseks saab lisada müra eelnevalt ettevalmistatud näidistest, et mudel saaks eraldada kasutaja kõne kõrvalistest heliallikatest. Kuid lihtsuse mõttes liidab autor selles töös algandmed juhuslike väärtustega [35].

4.2.4 One Hot Encoding

One Hot Encoding – masinõppe kasulik funktsioon, kuna mõned masinõppe algoritmid ei saa töötada otse kategooriliste andmetega.

Kategoorilised andmed on muutujatüübid, millel on mittearvulised väärtused. Seda tüüpi muutujaid nimetatakse ka nominaalseteks [36].

Klassifitseerimisalgoritmide rakendamisel tihti tuleb kategoorilised andmed teisendada numbrilisteks [36].

Teegi sklearn.preprocessing OneHotEncoderit kasutatakse kategooriliste muutujate teisendamiseks arvtunnusteks. Selles protsessis luuakse iga kategooria jaoks uus binaarne muutuja väärtustega 0 või 1, mille väärtus on 1, kui kategooria on antud andmepunktis olemas, ja 0 vastasel juhul [36].

4.3 Närvivõrkude mudelid

Tänapäeval on närvivõrgud võimelised paljuks: tuvastada objekte, parandada fotode kvaliteeti, diagnoosida haigusi. [35-38]

Käesolevas magistritöös närvivõrkude mudelid on kasutatud kõne emotsioonide klassifitseerimise ülesande lahendamiseks.

4.3.1 Närvivõrk

Tehisnärvivõrk on neuronite kogum, mis suhtlevad üksteisega. Nad on võimelised andmeid vastu võtma, töötlema ja looma. [37-40]

Närvivõrk on matemaatiline mudel, mis töötab samadel põhimõtetel nagu inimese aju. Närvivõrke ei saa programmeerida selle sõna tavalises tähenduses. Õigem on öelda, et nad treenitakse, ja see on traditsiooniliste algoritmide ees üks peamisi eeliseid [41].

Kui närvivõrku treenitakse, "näidatakse" talle andmeid, mille põhjal on vaja midagi ennustada, ja nende jaoks õigeid vastuseid - seda nimetatakse treeningandmestikuks. Infot peaks olema palju – arvatakse, et vähemalt kümme korda rohkem kui võrgus olevaid neuroneid [37].

Sisendneuronid võtavad vastu teavet, teisendavad seda ja edastavad. Teabe sisu töödeldakse automaatselt valemite abil ja teisendatakse matemaatilisteks kordajateks [37].

Igal neuronil on "kaal", mis arvutatakse spetsiaalsete algoritmide abil. See näitab, kui olulised on neuroni näidud kogu võrgu jaoks. Närvivõrgu treenimise ajal neuronite kaalud muutuvad ja tasakaalustuvad nende optimaalsete väärtuste leidmiseni [37].

Närvivõrgu väljund teisendatakse vastuseks, mis klassifitseerimise ülesande korral on iga klassi tõenäosus [37].

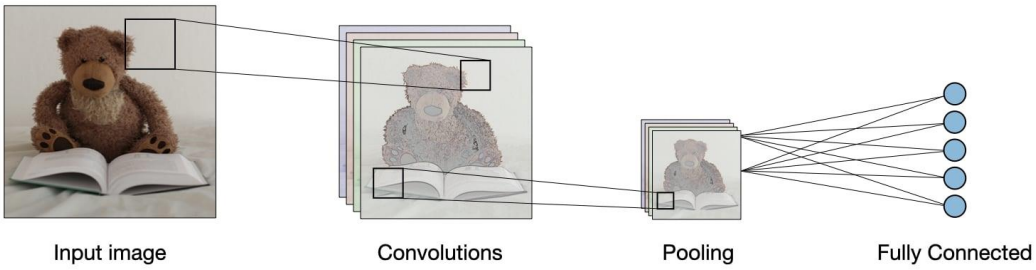
Tänapäeval on palju erinevaid närvivõrke tüüpe, mis on ettenähtud erinevat tüüpi probleemide lahendamiseks. Konvolutsioonilisi närvivõrke kasutatakse peamiselt multimeediaga seotud ülesannete lahendamisel: need töötavad graafika, heli ja videoga [40].

4.3.2 CNN

Konvolutsiooniliste närvivõrkude (*Convolutional Neural Network, CNN*) idee on konvolutsiooniliste kihtide vaheldus. Andmetüübi järgi jaotatakse kihid samamoodi nagu neuronid. Sisendkiht jaotab andmed järgmistele kihtidele ja väljundkiht moodustab lõpptulemuse. Andmeid töödeldakse sisemistes kihtides.

Mudeli ehitamiseks Kerases kasutatakse klassi Sequential, mis aktsepteerib kihtide loendit. Meetodi add() abil saab kihte lisada ka järjestikku [42].

Traditsiooniliste konvolutsiooniliste närvivõrkude arhitektuur koosneb üldiselt järgmistest kihtidest (Joonis 13):

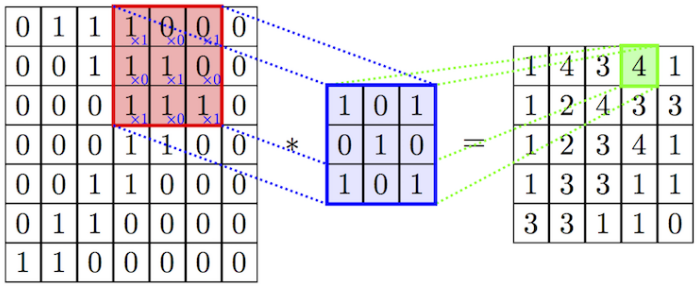


Joonis 13 Konvolutsioonilise võrgu kihid. [1]

Konvolutsiooniline kiht (*convolutional layer*, *CONV*) kasutab filtreid, mis teostavad konvolutsioonioperatsioone, skaneerides sisendandmeid nende mõõtmete järgi teatud sammuga. Selle kihi peamised hüperparameetrid on filtri suurus ja samm. Saadud väljundit nimetatakse tunnuste kaardiks [43].

[1] https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks

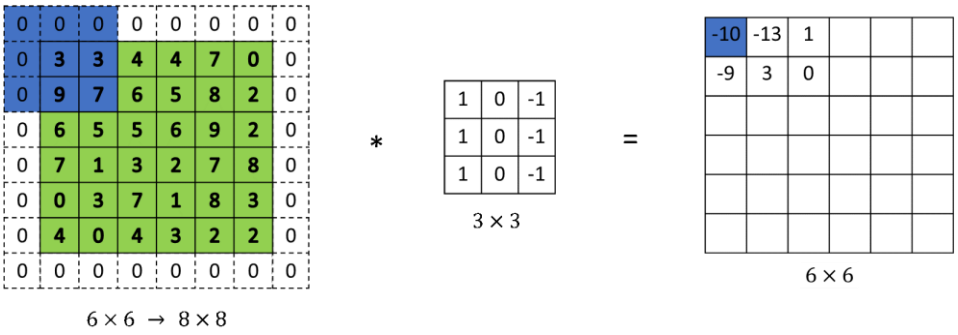
Joonis 14 näitab konvolutsiooni tehet, kus toimub sisendandmete elemendipõhine korrutamine filtriga ja tulemus summeeritakse. See summa kantakse tunnuste kaardile. Ja seega, liigutades filtrit mööda sisendandmeid, täidetakse tunnuste kaart.



Joonis 14. Konvolutsiooni operatsioon. [1]

Konvolutsioonikihi parameetrid [44]:

- Filtrite arv (*filters count, fc*) on kihis olevate filtrite arv.
- Filtri suurus (*filter size, fs*) on filtri kõrgus ja laius. Tavaliselt see on paaritu arv, kõige sagedamini kasutatavad filtrid suurusega 3x3 või 5x5.
- Konvolutsiooni samm (*stride, S*) on pikslite arv, mille võrra filtrimaatriks sisendpildil liigub. Mida suurem on samm, seda väiksemad on väljundis tunnuste kaardid.
- Ääris (*padding, P*) – on pikslite arv, mis lisatakse pildi igast servast. See väldib pildi vähendamist filtri suuruse võrra, kuna filtrit saab rakendada ainult nendes kohtades, kus iga filtri väärtuse all on pildi sisendväärtus. Joonis 15 näitab äärise lisamist.



Joonis 15. Ääris (*padding*).

[1] https://habr.com/ru/company/wunderfund/blog/314872/

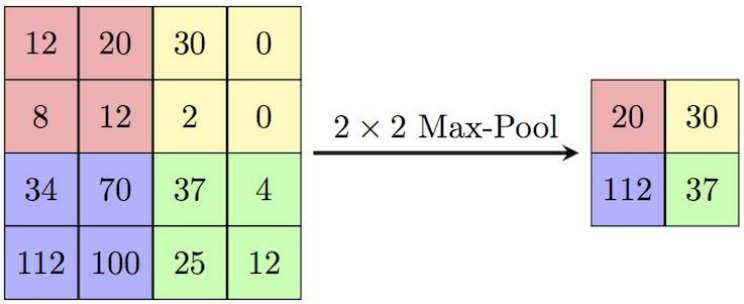
Konvolutsioonilise kihi väljundi suuruse arvutamiseks, kui konvolutsioonikihi sisendparameetrid suurusega $W_{in} \times H_{in} \times D_{in}$ on valemid (5, 6, 7), siis on kihi väljundparameetriks tensor suurusega $W_{out} \times H_{out} \times D_{out}$, kus [44]

$$W_{out} = \frac{W_{in} - fs + 2P}{S} + 1, \quad (5)$$

$$H_{out} = \frac{H_{in} - fs + 2P}{S} + 1, \quad (6)$$

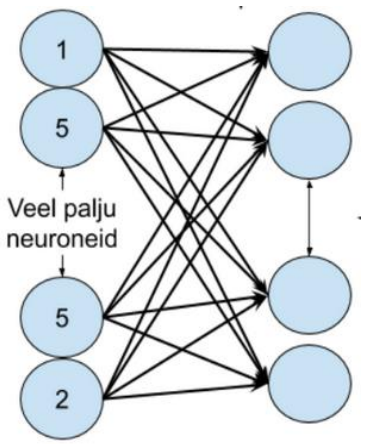
$$D_{out} = fc. \text{ [45]} \quad (7)$$

Ahenduskiht (*Pooling*) – operatsioon, mida tavaliselt rakendatakse pärast konvolutsioonikihti rakendamist tunnuste kaartide mõõtmete vähendamiseks ja olulise teabe eraldamiseks. Ahenduskihis rakendatakse filtreid samamoodi nagu konvolutsioonikihis, kuid konvolutsiooni asemel võetakse maksimaalne) või keskmine (*average pooling*) väärtus (Joonis 16) [43].



Joonis 16. Ahenduskihi rakendamine.

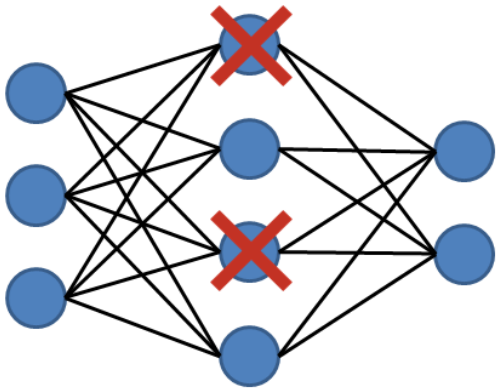
Täissidus kiht (*Dense / Fully Connected F*C) on kiht, mille väljundneuronid on ühendatud kõigi sisendneuronitega (Joonis 17) [46]. Kui need on olemas, asuvad need kihid tavaliselt CNN-i arhitektuuri lõpus. Täissidusad kihid on võimelised väga hästi õppima sisendandmete mittelineaarset kombinatsioone ja neid saab kasutada selliste eesmärkide optimeerimiseks nagu klassihinnangud [43].



Joonis 17. Täissidus kiht.[1]

Lamendavad (*Flatten*) kihid kasutatakse sisendandmete dimensionaalsuse vähendamiseks ehk lineaarseks vektoriks teisendamiseks.

Dropout-meetod (*Dropout*) kasutatakse närvivõrkudes ületreenimise probleemi lahendamiseks. Selleks valib ta juhuslikult neuronidest osa ja määrab neile väärtuse 0 iga kord, kui seda värskendatakse (Joonis 18) [47].



Joonis 18. Dropout-meetod. [2]

Ületreenimine on närvivõrgu liialt täpne sobitamine konkreetse treeningnäidete kogumiga, mille tulemusena kaob närvivõrgu üldistusvõime [48].

4.3.3 Mudeli kompileerimine

Kerases mudeli kompileerimiseks on meetod compile().

Peamine, mida kompileerimise käigus saab määrata, on kaofunktsioon.

Kaofunktsioon - arvutab kui hästi see mudel töötab, võrreldes mudel ennustusi andmepunktide tegelike väärtustega [49].

Rist-entroopia kaofunktsioon (*ingl Cross entropy loss function*) on funktsioon, mida kasutatakse binaarklassifitseerimise mudeli treenimisel [50].

[1] https://courses.cs.ut.ee/2020/Tehisintellekti_algkursus/Main/PARTIITehisn%C3%A4rviv%C3%B5rgud

[2] https://towardsdatascience.com/coding-neural-network-dropout-3095632d25ce

Kategoorilist ristentroopiat (*ingl* Categorical crossentropy) kasutatakse mudeli kaofunktsioonina, kui väljundina on vaja ennustada mitut klassi (Valem 8):

$$CE = -\sum_{i=1}^{i=N} y_{true_i} \cdot \log\left(y_{pred_i}\right)^{1} \quad (8)$$

kus y_true on andmeklassi tegelik ja y_pred on närvivõrgu poolt prognoositav väärtus. [51]

Teine oluline parameeter on optimeerimisalgoritm. Hästi valitud optimeerija võimaldab mudelit kiiremini treenida ja võimalusel vältida kohalikke miinimume.

Kõige polulaarsemad algoritmid on SGD, RMSprop, Adagrad, Adadelta, Adam, Adamax.

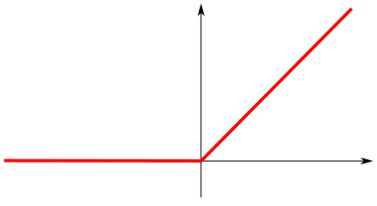
Adam (*ingl adaptive moment estimation*) on optimeerimisalgoritm, mis laiendab Adagradi ja RMSpropi, salvestades kaks gradiendi teabega loendit. See võimaldab parandada iga kaalu treeningkiiruse seadistust [52].

Adam meetodit loetakse hüperparameetrite väärtuste valiku suhtes üsna vastupidavaks ja seetõttu soovitatakse seda sageli vaikimisi meetodina [53]. Adam meetod on arvutuslikult tõhus, nõuab vähe mälu, invariantne gradientide diagonaalse skaleerimise suhtes ja see sobib hästi andmete/parameetrite poolest suurte probleemide lahendamiseks [54].

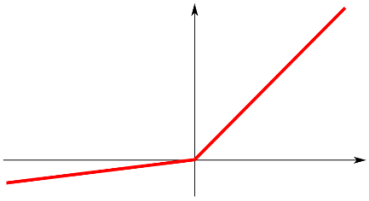
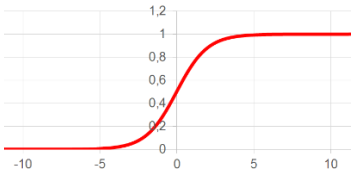
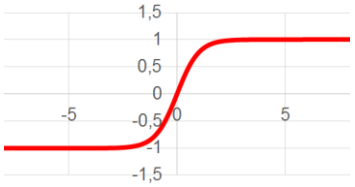
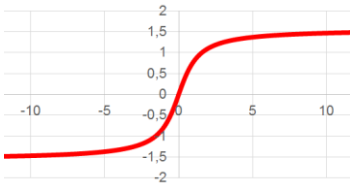
4.3.3.1 Aktiveerimisfunktsioon

Aktiveerimisfunktsioon on vajalik närvivõrgu mittelineaarsuse tagamiseks. Tabelis 2 on toodud mõned närvivõrkudes enim kasutatud aktiveerimisfunktsioonid [55].

Tabel 2. Neuronite aktiveerimise funktsioonid

№	Nimetus	Valem	Kujutis
1	*Rectified linear unit* (ReLU)	$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$	

[1] https://neuralthreads.medium.com/categorical-cross-entropy-loss-the-most-important-loss-function-d3792151d05b

№	Nimetus	Valem	Kujutis
2	Parameetriline *rectified linear unit* (PReLU)	$f(x) = \begin{cases} \alpha x, & x < 0 \\ x, & x \geq 0 \end{cases}$	
3	Logistiline funktsioon ehk sigmoid	$f(x) = \frac{1}{1 + e^{-x}}$	
4	Hüperboolne tangens (TanH)	$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$	
5	Arkustangens (ArcTan)	$f(x) = \arctan(x)$	

Selles töös kasutatakse ReLU aktiveerimisfunktsioone ja Softmax'i väljundkihi jaoks.

Aktiveerimisfunktsiooni ReLU loetakse praegu lihtsamaks ja arvutuslikult tõhusamaks [56].

Kui klassifitseerimisel on rohkem kui kaks klassi, kasutatakse närvivõrgu viimaseks kihiks Softmaxi kihti [57]. Funktsioon Softmax teisendab sisendväärtused klasside tõenäosusteks (Valem 9):

$$f_i(x) = \frac{exp(x_i)}{\sum_j exp(x_i)} \tag{9}$$

4.3.4 Mudeli treenimine

Närvivõrgu mudeli treenimiseks Kerases kasutatakse meetodit fit(), mis treenib mudelit kindlaksmääratud arvu epohhide (iteratsioonide) jooksul.

Epohh on treenimisprotsessis treenimisandmestiku üks täielik läbimine läbi närvivõrgu.

Kuna kogu andmestikku ei ole võimalik koheselt närvivõrku üle kanda, jagatakse andmestik mitmeks paketiks ehk osaks. Paketi suurus (*batch_size*) on ühes paketis pakutavate treenimisnäidete arv. [58]

Iteratsioonid on ühe epohhi läbimiseks vajalik paketide arv.

Sellel kujul saab mudelit treenida, kuid seda saab teha palju tõhusamalt, kui kasutate tagasikutsefunktsioonid (callbacks). Neid saab kasutada treenimise varaseks lõpetamiseks, et vältida ületreenimist.

ReduceLROnPlateau tagasikutset saab kasutada õppimiskiiruse vähendamiseks, kui valideerimisandmete kadu enam ei vähene. Õppimiskiiruse vähendamine või suurendamine kaotuskõvera käänupunktis on tõhus strateegia kohalikust miinimumist väljumiseks treeningu ajal [59].

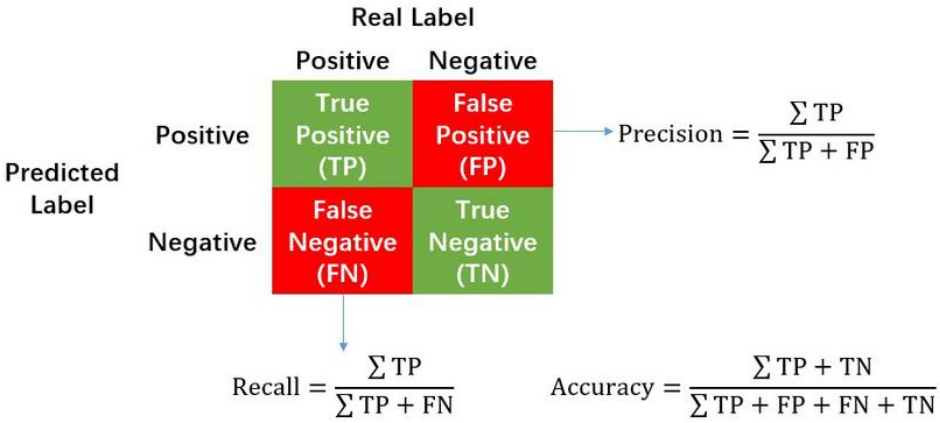
EarlyStopping tagasikutset saab kasutada treeningprotsessi katkestamiseks, kui kasutatav sihtmõõdik ei ole teatud perioodide jooksul paranenud. Seda tagasihelistamist kasutatakse tavaliselt koos ModelCheckpoint tagasikutsega, mis võimaldab treeningu ajal salvestada mudeli oleku. Võib kasutada ka CSVLoggerit, mis logib epohhi tulemused csv-faili.

Mudeli treenimise väljund näitab iga epohhi täitmise tulemust treenimis- ja valideerimisandmetel. Kuvatakse epohhi number, pakettide arv, kulunud aeg ja vead, arvutatud kaofunktsioon ja mõõdikud.

Treenitud mudeli testandmetel hindamiseks kasutab Keras evaluate() meetodit. Prognoositud väärtuste saamiseks on predict() meetod. Meetodi classification_report abil koostatakse tekstiaruanne, mis näitab peamisi mõõdikuid mudeli hindamiseks.

4.3.5 Mõõdikud mudeli hindamiseks

Peamine mõõdik klassifitseerimismudeli täpsuse hindamiseks on segadusmaatriks. See mõõdik annab ülevaate sellest, kui hästi mudel töötab. Joonis 19 Joonis 19. näitab enamikku mõõdikutest, mida saab segadusmaatriksist tuletada.



Joonis 19. Täpsuse, korrektsuse ja õigsuse arvutamine [1]

Õigsus (*ingl. accuracy*) on õigete ennustuste koguarv jagatud ennustuste koguarvuga.

Täpsus (*ingl. precision*) määrab, kui usaldusväärne on tulemus, kui mudel vastab, et punkt kuulub positiivse klassi. [60]

Saagis = korrektsus (*ingl. recall*) väljendab, kui hästi mudel suudab positiivse klassi tuvastada. [60]

F1 skoor (*ingl. F1 score*) ühendab täpsuse ja korrektsuse väärtused ühes mõõdikus, arvutades nende harmoonilist keskmist (Valem 10) [60]:

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} \qquad (10)$$

1

https://www.researchgate.net/publication/336402347_Analyzing_the_Leading_Causes_of_Traffic_Fatalities_Using_XGBoost_and_Grid-Based_Analysis_A_City_Management_Perspective

5 Töö tulemused

Selles jaotises esitatakse mudeli treenimise parameetreid ja peamisi mõõdikuid. Eksperimente viidi läbi jaotises 4 kirjeldatud tööriistade abil. Mudelite täpsust hinnati kasutades segadusmaatriksit, õigsust, täpsust, korrektsus ja F1 skoori.

5.1 Mudel

5.1.1 Andmete ettevalmistamine

Mudeli treenimiseks valis autor pärast muudetud andmete kuulamist andmete rikastamiseks müra ja helikõrguse. Andmete rikastamiseks kasutati müra lisamist mooduli NumPy random.uniform funktsiooni abil ja helikõrguse muutmist mooduli librosa pitch_shift funktsiooni abil.

Mudeli treenimiseks valiti emotsioonide tuvastamisel sageli kasutatavad tunnused, mida saab librosa teegi abil ekstraheerida: MFCC, RMSE, Zero-Crossing Rate. Teegi funktsioon saab failitee ja laadib helifaili tunnuste ekstraktimiseks, mis seejärel koondatakse ja tagastatakse numpy massiivina. Tunnuste arv sai 2376.

Järgmisena teisendati OneHotEncoderi abil emotsioonide nimetused numbriliseks.

Pärast kõigi failide individuaalset töötlemist ja tunnuste ekstraheerimist, jagati andmekogum proportsioonis 85-11-4 kolmeks: treening, test ja valideerimis osaks.

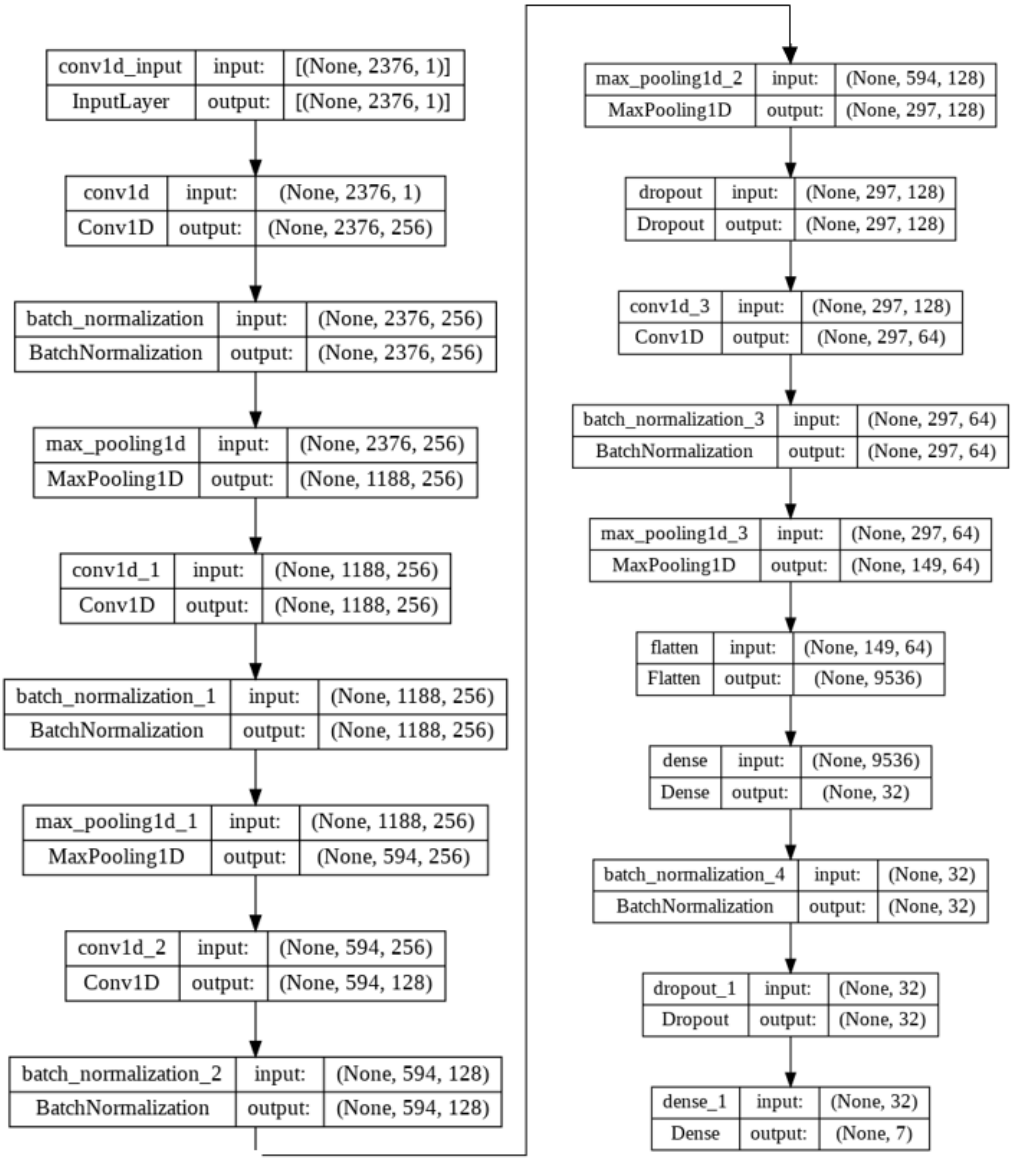
Tulemusena treeningandmestik sisaldas 42894 näidist, valideerimisandmestik - 1893 näidist ja testandmestik - 5677 näidist.

5.1.2 Mudeli arhitektuur

Kõik mudeli olid treenitud Keras paketi abil.

Autor testis mitmeid närvivõrkude arhitektuure: mitmekihilised närvivõrgud, konvolutsioonilised närvivõrgud (CNN), rekurrentsed närvivõrgud (LSTM). Samuti proovis kasutada erinevad tunnused ja rikastamata andmed. Mõned tulemused on Lisas 3.

Töös kirjeldatud mõõdikute kohaselt saavutati parimad tulemused CNN algoritmi kasutamisel järgneva arhitektuuriga (vt Joonis 20).



Joonis 20. Mudeli arhitektuur

Parimas leitud CNN mudelis on kuus kihti – neli Conv1D kihti ja kaks täissidus kihti (Dense), et vältida ületreenimist. Dimensioonide vähendamiseks kasutati ka ahenduskihte (MaxPooling1D).

Mudel treeniti hüperparameetritega `batch_size=128`, `epochs=100` ning eelkatkestusega 67 iteratsioonil. Optimeerimismeetoditest kasutati Adam algoritmi, aktiveerimisfunktsiooniks igal peidetud kihil oli ReLU, väljundkihiks kasutati Softmax funktsiooni. Õppimiskiiruse alumine piir learning_rate= 0.000015.

Mudeli õpe peatükis 4.1 toodud omadustega arvutis võttis aega ca 8 tundi. Lisas 2 on lingid Colab'ile, kus on näha treenimise tulemused.

5.1.3 Mudeli hinnang

Parima CNN mudeli täpsuse näitajad on järgmised: treenimisandmete kadu on 0.0788 ja õigsus on 0.9723, valideerimisandmete kadu on 0.2168 ja õigsus on 0.9408.

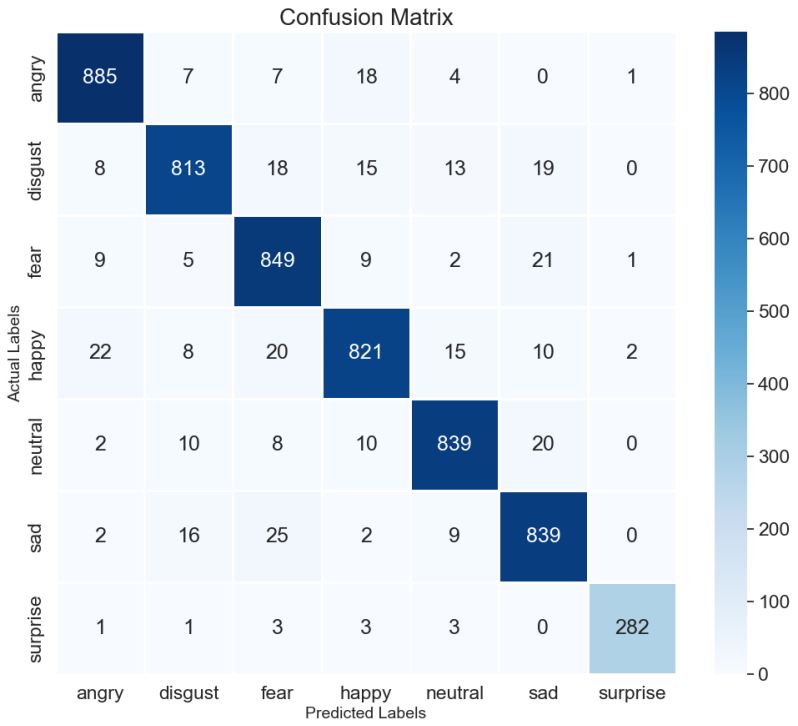
Järgmistel piltidel (Joonis 21, Joonis 22) on visualiseeritud mudeli treenimise prosess. Joonis 21 näitab mudeli treenimisajal kaofunktsiooni graafiku ja Joonis 22 näitab mudeli treenimisajal õigsuse graafiku. Ja enne kui toimus ületreenimine protsess oli lõppenud.

Joonis 21 Treenimis ja valideerimisandmete kaofunktsiooni graafik

Joonis 22 Treenimis ja valideerimisandmete õigsuse graafik

Edasi treenitud mudeli hindamiseks oli kasutatud testimisandmeid. Testandmetel õigsus oli 0.9385 ja kadu oli 0.1863. See näitab, et mudeli täpsus testandmetel on 93.85 %.

Segadusmaatriks (Joonis 23) samuti näitab, et tulemused on head, tegelike ja mudeli poolt ennustatud andmete võrdlemisel.



Joonis 23. Segadusmaatriks.

Tekstiaruandes (Tabel 3) on näha mudeli hindamise peamised mõõdikud ning igal emotsiooniklassil on kõrge täpsus, korrektsus ja F1 skoor. Nagu on näha näitajate järgi on kõige paremini tuvastatud emotsioonid on "viha" ja "üllatus".

Tabel 3. Peamised mõõdikud mudeli hindamiseks

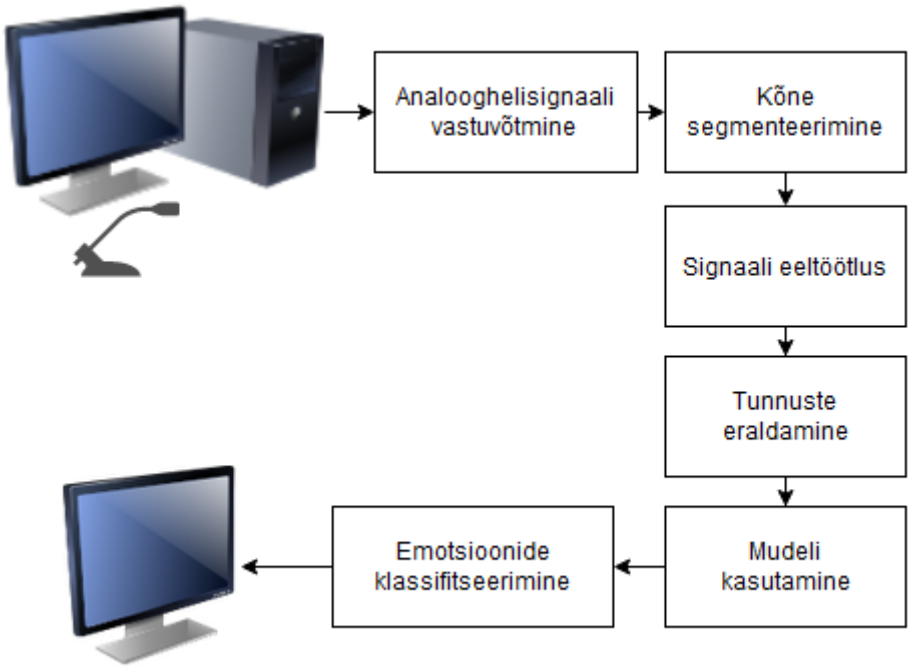
	precision	**recall**	**f1-score**	**support**
angry	**0.95**	**0.96**	**0.96**	922
disgust	**0.95**	0.92	0.93	886
fear	0.91	0.95	0.93	896
happy	0.94	0.91	0.92	898
neutral	0.95	0.94	0.95	889
sad	0.92	0.94	0.93	893
surprise	**0.99**	**0.96**	**0.97**	293
accuracy			0.94	5677
macro avg	0.94	0.94	0.94	5677
weighted avg	0.94	0.94	0.94	5677
weighted avg	0.93	0.93	0.93	5473

5.2 Veebirakenduse prototüüp

Käesoleva töö tulemusena leitud kõne emotsioonide tuvastamise mudel oli kasutatud veebirakenduse loomiseksplõpilaste ja õppejõudude emotsionaalse seisundi jälgimiseks õppeprotsessis.

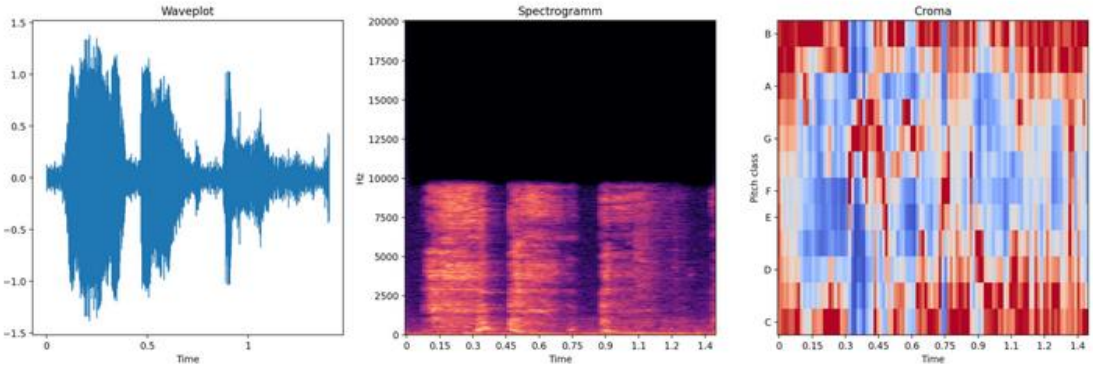
Emotsioonituvastussüsteemi töö saab kirjeldada järgmise sammude jadaga: analooghelisignaali vastuvõtmine sisendina, kõne segmenteerimine, signaali eeltöötlus, mille käigus iga kõnesegment teisendatakse tunnusvektoriks ning emotsioonide klassifitseerimine loodud mudeli abil (Joonis 24).

Kõne segmenteerimise all mõistetakse tavaliselt kõnevoo jagamist mõneks elemendiks - foneemideks, silpideks, sõnadeks [62].



Joonis 24. Rakenduse skeem

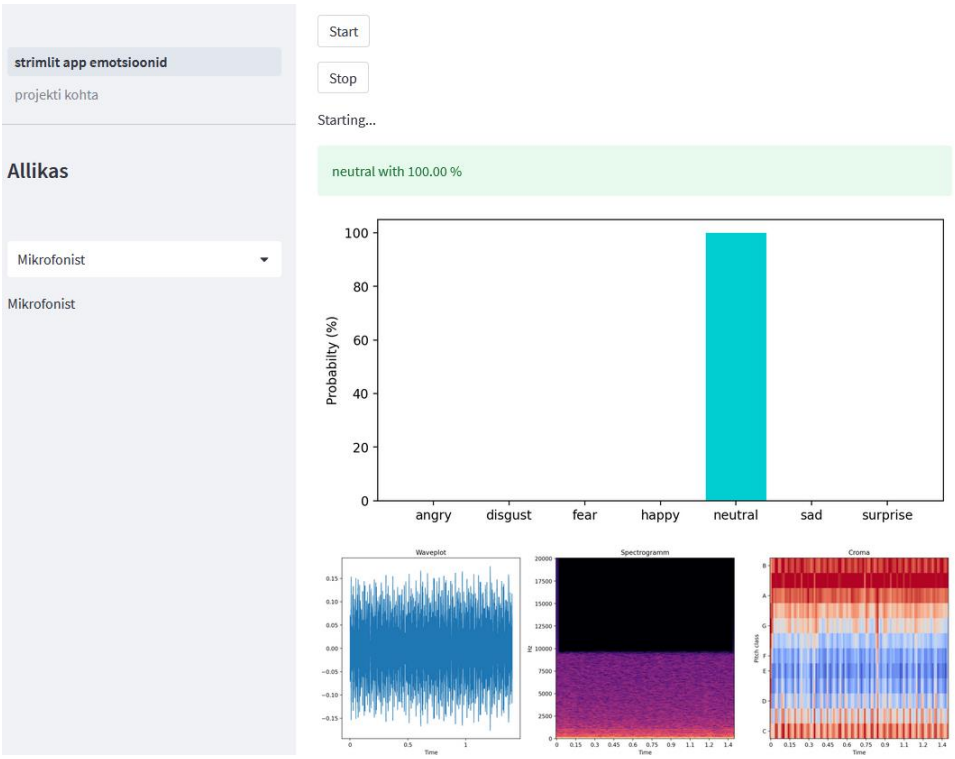
Analoogsignaali vastuvõtmine toimub arvuti mikrofooni abil. Saadud fail salvestatakse arvutis. Kuna autor kasutas kahesekundilist salvestamist, siis fail kohe saadetakse signaali eeltöötluseks. Faili töödeldakse sama funktsioonidega, mida kasutati mudeli loomisel, et saadud tunnuste lõplik mõõde langes kokku mõõtmetega, mida mudel sisendis ootab. Parast tunnuste eraldamist, rakendatakse mudelit emotsiooni klassifitseerimiseks. Klassifitseerimise tulemused näidetakse protsentides tulba diagrammi abil. Audiosignaal visualiseerimiseks kasutatakse lainegraafikut, spektrogrammi ja kromagrammi (Joonis 25).



Joonis 25. Veebirakeduses audiosignaali visualiseerimine.

Rakendus oli kirjutatud Python programmeerimis keele kasutades raamistikku Streamlit.

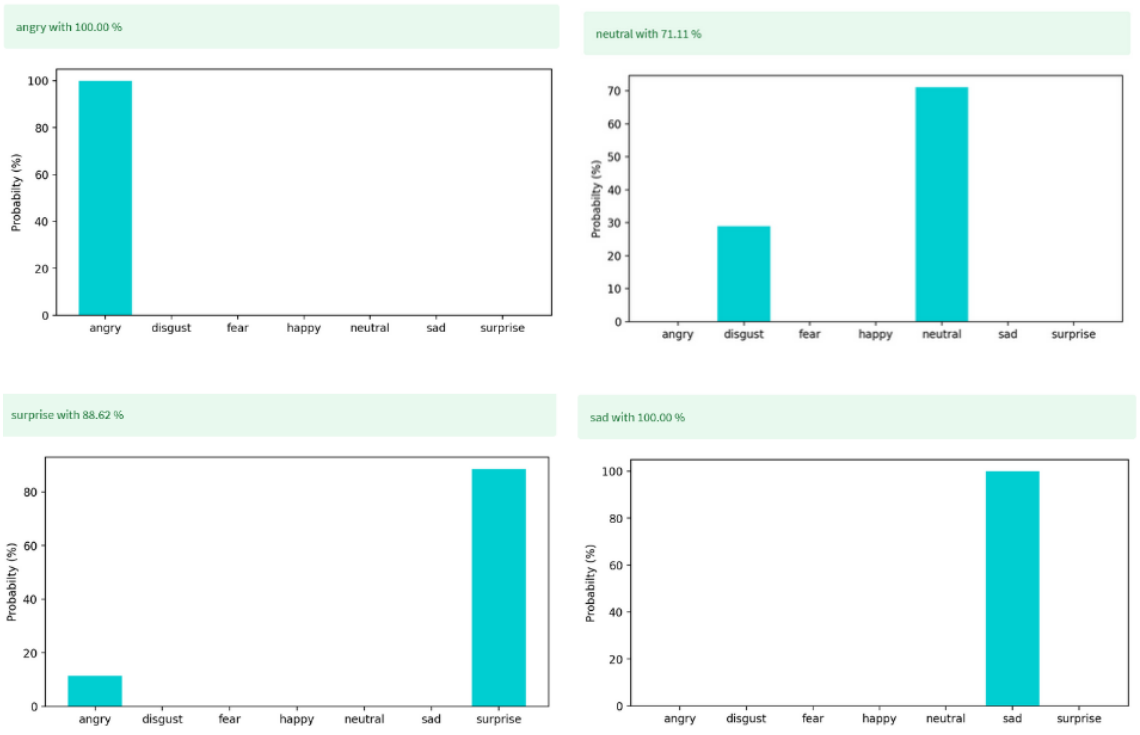
Rakenduse vaade emotsiooni mikrofoonist tuvastamisega on esitatud Joonis 26 ja audiofailist tuvastamisega on esitatud Lisas 3.



Joonis 26. Veebirakenduse vaade

Antud etapil on loodud veebirakenduse prototüüp, mis ei ole veel reaalse või simuleeritud õppeprotsessis katsetatud ning selle funktsionaalsus nõuab veel täiendamist nagu kasutajate autentimise, andmete andmebaasi salvestamise, õppeprotsesse emotsionaalse tausta visualiseerimise ja analüüsimise lisamist. Audiofailide andmebaasi salvestamise lisamine annab võimaluse edasi uute näidistega täiendatud andmetel mudeli üle treenida ja selle kvaliteedi parandada.

Autor ise katsetas mudeli toimimist (Joonis 27) ning testimise tulemused näitasid, et rakendus lubab reaalajas kõne järgi emotsioone tuvastada. Kuid emotsioonide tuvastamise täpsuse mõttes selle katsetamise tulemused on subjektiivsed ja mudel nõuab veel katsetamist suurema hulga kasutajate poolt.



Joonis 27. Mudeli töö testimine.

Edasiarendamisel planeeritakse kasutada automaatne hääletegevuse tuvastamine (*Voice Activity Detection*, VAD). Emotsioonide tuvastamise kvaliteedi tõstmiseks on võimalik luua ja ühendada käesoleva mudeliga näo emotsioonide tuvastamise mudeli. Edasi võib katsetada ka mudeli, mis eristab ainult positiivseid, negatiivseid ja neutraalseid emotsioone. Ning võib proovida luua emotsioonide tuvastamise mudeli, jagades kõneleja soo järgi.

6 Kokkuvõte

Selle magistritöö eesmärk oli luua meetod õppeprotsessi käigus üliõpilaste ja õppejõudude emotsionaalse seisundi jälgimiseks kõne järgi süvaõppe meetodite abil.

Töös püstitatud ülesanded on täidetud. On käsitletud teoreetiline osa ja olemasolevaid lahendusi antud valdkonnas. On uuritud erinevaid artikleid süvanärvivõrkude treenimise ja emotsioonide tuvastamise kohta s.h. seotud emotsioonide tuvastamisega õppeprotsessis.

Kasutades süvaõppe algoritme on leitud CNN närvivõrgumudel, mille emotsioonide tuvastamise täpsus testimisandmetel on 93.85%. Saadud CNN-i mudeli põhjal on loodud emotsioonide tuvastamise rakenduse prototüüp, mis võimaldab üliõpilaste ja õppejõudude emotsionaalse seisundi reaalajas jälgida. Loodud veebirakendus on edasi plaanis õppeprotsessi või selle simulatsiooni tingimustes katsetada ja selle funktsionaalsust täiendada.

Tulevikus autoril on plaanis ka mudeli efektiivsust reaalsetes tingimustes testida ja parandada. Näiteks, võib uute näidistega täiendatud andmetel mudeli üle treenida, või proovida mudeli, mis eristab ainult positiivseid, negatiivseid ja neutraalseid emotsioone. Ning võib proovida luua emotsioonide tuvastamise mudeli, jagades kõneleja soo järgi.

Tulevikus on ka soov ühendada kahte mudelit: kõnes emotsioonide tuvastamise koos näoemotsiooni tuvastamisega.

Edasiarendamisel planeeritakse kasutada automaatne hääletegevuse tuvastamine (*Voice Activity Detection*, VAD) ja andmebaasi kasutamist, et salvestada kõik andmed edaspidiseks analüüsimiseks.

Kasutatud kirjandus

1. Russell, James. (1980). A Circumplex Model of Affect. Journal of Personality and Social Psychology. 39. 1161-1178. 10.1037/h0077714.
2. El Ayadi M, Kamel M S, Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases [J]. Pattern Recognition, 2011, 44(3): 572-587.
3. Reza Chu. Speech Emotion Recognition with Convolutional Neural Network. https://towardsdatascience.com/speech-emotion-recognition-with-convolution-neural-network-1e6bb7130ce3
4. Husbaan I. Attar, Nilesh K. Kadole. Speech Emotion Recognition System Using Machine Learning. https://ijrpr.com/uploads/V3ISSUE5/IJRPR4210.pdf
5. Cen, Ling & Wu, Fei & Yu, Zhu & Hu, Fengye. (2016). A Real-Time Speech Emotion Recognition System and its Application in Online Learning. https://www.researchgate.net/publication/313139706_A_Real-Time_Speech_Emotion_Recognition_System_and_its_Application_in_Online_Learning
6. R. Anusha, P. Subhashini, D. Jyothi, P. Harshitha, J. Sushma and N. Mukesh, "Speech Emotion Recognition using Machine Learning," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 1608-1612.
7. Chen Jin, A.I. Sherstneva, I.A. Botygin. Speech Emotion Recognition Based On Deep Residual Convolutional Neural Network. 2022. https://journalpro.ru/pdf-article/?id=14383
8. L. Jie, Z. Xiaoyan and Z. Zhaohui, "Speech Emotion Recognition of Teachers in Classroom Teaching," 2020 Chinese Control And Decision Conference (CCDC), 2020, pp. 5045-5050.
9. Ristea, Nicolae-Catalin & Dutu, Liviu-Cristian & Radoi, Anamaria. (2020). Emotion Recognition System from Speech and Visual Information based on Convolutional Neural Networks. https://www.researchgate.net/publication/339642462_Emotion_Recognition_System_from_Speech_and_Visual_Information_based_on_Convolutional_Neural_Networks

10. В.В. Видман, А.Я. Видман, В.М. Саклаков. (2018) Распознавание эмоций из речевого сигнала. https://earchive.tpu.ru/bitstream/11683/52680/1/conference_tpu-2018-C04_p140-141.pdf
11. Д.Суворов.40 проектов на Python для новичков и продвинутых разработчиков [Võrgumaterjal]. Saadaval: https://proglib.io/p/40-proektov-na-python-dlya-novichkov-i-prodvinutyh-razrabotchikov-2022-05-13. [Kasutatud 10.2022].
12. Что такое Numpy? https://stepik.org/lesson/241329/step/1?unit=213910
13. 9. aasta 2021 parimat Pythoni teeki masinõppe jaoks [Võrgumaterjal]. Saadaval: https://www.uniquenewsonline.com/et/9.-aasta-2020-parimat-pythoni-teeki-masin%C3%B5ppe-jaoks/amp/ . [Kasutatud 10.2022].
14. Valter Kiisk. Masinõpe. Scikit-Learn ja TensorFlow [Võrgumaterjal]. Saadaval: https://kodu.ut.ee/~kiisk/python/machine.html#Regressioon-(Scikit-Learn)
15. Andrus Rinde. Multimeedium, digitaalsed helisalvestused. Helid. http://www.cs.tlu.ee/~rinde/mm_materjal/pdf/mm_audio.pdf
16. Perioodilised lained [Võrgumaterjal]. Saadaval: https://opik.fyysika.ee/index.php/book/section/3693. [Kasutatud 10.2022].
17. Sagedus [Võrgumaterjal]. Saadaval: https://et.wikipedia.org/wiki/Sagedus
18. Александр Радзишевский, Александр Чижов. Цифровой звук [Võrgumaterjal]. Saadaval: https://studfile.net/preview/6809251/page:4/
19. Andrus Rinde. Multimeedium, digitaalsed helisalvestused. Helid arvutis. http://www.cs.tlu.ee/~rinde/mm_materjal/pdf/mm_audio_wave.pdf
20. Кодирование звуковой информации [Võrgumaterjal]. Saadaval: https://sites.google.com/site/informatika5796/dvuazycnye/kodirovanie-zvukovoj-informacii . [Kasutatud 10.2022].
21. Кодирование звуковой информации [Võrgumaterjal]. Saadaval: https://www.bsuir.by/m/12_104571_1_116067.pdf . [Kasutatud 10.2022].
22. В.А. Волохов, О.В. Махныткина, И.Д. Мещеряков. Методические указания к выполнению лабораторных работ по курсу „цифровая обработка сигналов“ . https://books.ifmo.ru/file/pdf/3111.pdf .
23. Tushar Gupta. Speech Emotion Detection [Võrgumaterjal]. Saadaval: https://medium.com/@tushar.gupta_47854/speech-emotion-detection-74337966cf2 . [Kasutatud 11.2022].

24. Введение в библиотеку librosa [Võrgumaterjal]. Saadaval: https://pythonru.com/biblioteki/librosa. [Kasutatud 10.2022].
25. Mel Frequency Cepstral Coefficient (MFCC) tutorial [Võrgumaterjal]. Saadaval: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs . [Kasutatud 10.2022].
26. А.Хайбуллин. Анализ аудиоданных с помощью глубокого обучения и Python (часть 1) [Võrgumaterjal]. Saadaval: https://medium.com/nuances-of-programming/анализ-аудиоданных-с-помощью-глубокого-обучения-и-python-часть-1-2056fef8525eh . [Kasutatud 11.2022].
27. Comparison of the RMS Energy and the Amplitude Envelope [Võrgumaterjal]. Saadaval: https://www.analyticsvidhya.com/blog/2022/05/comparison-of-the-rms-energy-and-the-amplitude-envelope/
28. Maël Fabien. Sound Feature Extraction [Võrgumaterjal]. Saadaval: https://maelfabien.github.io/machinelearning/Speech9/. [Kasutatud 11.2022].R. Pappagari, J. Villalba, P. Żelasko, L. Moro-Velazquez and N. Dehak, "CopyPaste: An Augmentation Method for Speech Emotion Recognition," *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 6324-6328.
29. Как делать аугментацию данных для задачи распознавания объектов [Võrgumaterjal]. Saadaval: https://neurohive.io/ru/novosti/kak-delat-augmentaciju-dannyh-dlya-raspoznavaniya-obektov/ . [Kasutatud 11.2022].
30. Edward Ma. Data Augmentation for Audio [Võrgumaterjal]. Saadaval: https://medium.com/@makcedward/data-augmentation-for-audio-76912b01fdf6
31. Ristea, Nicolae-Catalin & Dutu, Liviu-Cristian & Radoi, Anamaria. (2020). Emotion Recognition System from Speech and Visual Information based on Convolutional Neural Networks. https://www.researchgate.net/publication/339642462_Emotion_Recognition_System_from_Speech_and_Visual_Information_based_on_Convolutional_Neural_Networks
32. Учебник Cubase. Pitch Shift. Saadaval: https://cubase.su/publ/1-1-0-346
33. Keyur Paralkar. Audio Data Augmentation in python [Võrgumaterjal]. Saadaval: https://medium.com/@keur.plkar/audio-data-augmentation-in-python-a91600613e47 . [Kasutatud 11.2022].

34. N. Braunschweiler, R. Doddipatla, S. Keizer and S. Stoyanchev, "A Study on Cross-Corpus Speech Emotion Recognition and Data Augmentation," 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2021, pp. 24-30.
35. A. Mujaddidurrahman, F. Ernawan, A. Wibowo, E. A. Sarwoko, A. Sugiharto and M. D. R. Wahyudi, "Speech Emotion Recognition Using 2D-CNN with Data Augmentation," 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM), 2021, pp. 685-689.
36. Как выполнить горячее кодирование данных в Python [Võrgumaterjal]. Saadaval: https://pythonpip.ru/osnovy/one-hot-encoding-python. [Kasutatud 11.2022].
37. Как работает нейронная сеть: разбираемся с основами [Võrgumaterjal]. Saadaval: https://blog.skillfactory.ru/kak-rabotaet-nejronnaya-set-razbiraemsya-s-osnovami/ . [Kasutatud 11.2022].
38. Нейросети: как появились, зачем нужны и чего от них ждать [Võrgumaterjal]. Saadaval: https://blog.sibirix.ru/neural_networks/ . [Kasutatud 11.2022].
39. Нейросети [Võrgumaterjal]. Saadaval: https://www.tadviser.ru/index.php/Статья:Нейросети_(нейронные_сети). [Kasutatud 11.2022].
40. Нейронные сети: какие бывают и как их используют бренды [Võrgumaterjal]. Saadaval: https://blog.ingate.ru/detail/neyronnye-seti-kakie-byvayut-i-kak-ikh-ispolzuyut-brendy/ . [Kasutatud 11.2022].
41. Нейросети [Võrgumaterjal]. Saadaval: https://hi-news.ru/tag/nejroseti. [Kasutatud 11.2022].
42. Г.Бузмарев. Первое знакомство с нейронными сетями на примере Tensorflow 2 [Võrgumaterjal]. Saadaval: https://tproger.ru/articles/pervoe-znakomstvo-s-nejronnymi-setjami-na-primere-tensorflow-2/.
43. Shervine Amid. Convolutional Neural Networks cheatsheet [Võrgumaterjal]. Saadaval: https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks . [Kasutatud 12.2022].
44. Перминов А. Свёрточная нейронная сеть с нуля. Часть 0 [Võrgumaterjal]. Saadaval: https://programforyou.ru/poleznoe/convolutional-network-from-scratch-part-zero-introduction

45. V.Dumoulin and F.Visin. A guide to convolution arithmetic for deep learning [Võrgumaterjal]. Saadaval: https://arxiv.org/pdf/1603.07285.pdf. [Kasutatud 11.2022].
46. Перминов А. Свёрточная нейронная сеть с нуля. Часть 4. Полносвязный слой [Võrgumaterjal]. Saadaval: https://programforyou.ru/poleznoe/convolutional-network-from-scratch-part-four-fully-connected-layer. [Kasutatud 12.2022].
47. Слои Keras: параметры и свойства / keras 5 [Võrgumaterjal]. Saadaval: https://pythonru.com/biblioteki/sloi-keras-parametry-i-svojstva-keras-5 . [Kasutatud 12.2022].
48. Переобучение - что это и как этого избежать, критерии останова обучения [Võrgumaterjal]. https://proproprogs.ru/neural_network/pereobuchenie-chto-eto-i-kak-etogo-izbezhat-kriterii-ostanova-obucheniya
49. Функция потерь (Loss Function) [Võrgumaterjal]. Saadaval: https://www.helenkapatsa.ru/funktsiia-potieri/ . [Kasutatud 11.2022].
50. A.Kumar. Keras – Categorical Cross Entropy Loss Function [Võrgumaterjal]. Saadaval: https://vitalflux.com/keras-categorical-cross-entropy-loss-function/
51. Е.С. Попова, В.Г. Спицын, Ю.А. Болотова. Программа и алгоритм сегментации и распознавания рукопечатных символов с помощью сверточных нейронных сетей. (2018) https://www.graphicon.ru/html/2018/papers/230-233.pdf
52. Эндрю Гласснер. Глубокое обучение без математики. Том 2. Практика. 2020. pp.535
53. Diederik P. Kingma, Jimmy Lei Ba. ADAM: a method for stochastic optimization. (2015) https://arxiv.org/pdf/1412.6980.pdf
54. Categorical cross-entropy loss — The most important loss function. https://neuralthreads.medium.com/categorical-cross-entropy-loss-the-most-important-loss-function-d3792151d05b
55. Введение в машинное обучение и искусственные нейронные сети [Võrgumaterjal]. Saadaval: https://foobar167.github.io/page/vvedeniye-v-mashinnoye-obucheniye-i-iskusstvennyye-neyronnyye-seti.html. [Kasutatud 12.2022].

56. Зачем в науке о данных нужны теория вероятностей и статистика [Võrgumaterjal]. Saadaval: https://remont-komp.ru/zachem-v-nauke-o-dannyh-nuzhny-teorija/. [Kasutatud 12.2022].
57. Softmax Regression Explained with Python Example. https://vitalflux.com/what-softmax-function-why-needed-machine-learning/
58. Эпоха против размера партии против итераций [Võrgumaterjal]. Saadaval: https://machinelearningmastery.ru/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9/. [Kasutatud 12.2022].
59. Мониторинг моделей глубокого обучения средствами библиотеки Keras. http://se.moevm.info/lib/exe/fetch.php/courses:artificial_neural_networks:pr_7.pdf
60. Метрики качества [Võrgumaterjal]. Saadaval: https://vickynomica.com/accuracy-recall-precision/. [Kasutatud 11.2022].
61. MA, Jun & Ding, Yuexiong & Cheng, Jack & Tan, Yi & Gan, Vincent & ZHANG, Jingcheng. (2019). Analyzing the Leading Causes of Traffic Fatalities Using XGBoost and Grid-Based Analysis: A City Management Perspective. IEEE Access. PP. 1-1.
62. С. В. Омельченко. Алгоритмы сегментации речевого сигнала на фоне коррелированной помехи. (2018) https://www.researchgate.net/publication/324957905_Algorithms_of_segmentation_of_speech_signal_on_the_correlated_noise_background/fulltext/5aed07d1a6fdcc8508b7f5b9/Algorithms-of-segmentation-of-speech-signal-on-the-correlated-noise-background.pdf

Lisa 1 – Lihtlitsents Lõputöö Reprodutseerimiseks Ja Lõputöö Üldsusele Kättesaadavaks Tegemiseks[1]

Mina, Valeria Juštšenko

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Süvaõppel põhinev kõne emotsioonide tuvastamine õppeprotsessis“, mille juhendaja on Olga Dunajeva
   1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
   1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

06.01.23

1 Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingulise tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.

Lisa 2 – Colab lingid

Andmestikute analüüs:

https://colab.research.google.com/drive/1gl8YxLEjFbJxVzWz5LB9ZKWYesyHwm_H?usp=sharing

Tunnuste eraldamine:

https://colab.research.google.com/drive/1Dtx32UXygJzE4-DROlua45bHeY9BxadX?usp=sharing

Mudeli treenimine:

https://colab.research.google.com/drive/1-ESJEcsY4jgbw29NO3vF6B6Knerf0gDB?usp=sharing

Lisa 3 – Mudelite tulemuste võrdlemine (5 andmestikutel).

Mudel	**LSTM**	**CNN Conv1 (nagu lõputöös kasutatud) ilma rikastamist**	**CNN Conv2**	**MLPClassifier**
Tunnused	mfcc(n_mfcc[1]= 13, n_fft[2]=2048)	MFCC, zcr, rmse,	mfcc(fmin=50, n_mfcc=30)	Keskmised väärtused: chroma_stft, mfcc, melspectrogram
Optimeerimis algoritm	Adam	Adam	Adam	Adam
Epohhid	44	79	89	1000
Accuracy testandmetel	45,77 %	59,51 %	61.17%	61.72%
Treenimise aeg	13 min	1 t 11 min	4 min	3 min

[1] Mel-sageduse kepstri kordajate arv
[2] sagedusribade arv

Lisa 4 – Rakenduse 2. leht. Tuvastamine failist