

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Natalja Ivleva 204415IABM

Masinõppel põhinev näo 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: Natalja Ivleva

28.12.2022

Annotatsioon

Käesoleva lõputöö eesmärgiks on luua masinõppel põhinev meetod üliõpilaste ja õppejõudude emotsionaalse seisundi jälgimiseks õppeprotsessi käigus näo emotsioonide järgi. Mudelite loomiseks on töös kasutatud nelja erinevat masinõppe meetodit: SVM, Random Forest, CNN ja VGG16 sügav konvolutsiooniline närvivõrk.

Masinõppe mudelid loodi Pythoni Scikit-Learn teegi abil, CNN algoritmid realiseeriti TensorFlow tagarakendusel töötavas Kerases, kõik eksperimendid teostati Google Colabis.

Mudelite treenimiseks antud töös kasutatakse andmestikku FER2013. Valesti märgendatud piltide kustutamiseks ja andmekogumi tasakaalustamiseks kasutati Deepface'i raamistikku.

Käesolevas magistritöös on käsitletud kolme tüüpi emotsioonide klassifitseerimist:

1. Klassifitseerimine 7 klassiks: üllatus, viha, hirm, kurbus, neutraalne, rõõm, vastikus.
2. Klassifitseerimine 2 klassiks: positiivsed (rõõm) ja negatiivsed (viha, hirm, kurbus, vastikus) emotsioonid.
3. Klassifitseerimine 2 klassiks: aktiivsed (viha, vastikus, hirm, rõõm, üllatus) ja passiivsed (kurbus) emotsioonid.

Parim tuvastustäpsus saavutati VGG16 algoritmiga (10-kordse ristvalideerimise keskmine täpsus 96%) positiivse/negatiivse klassifitseerimismudeli korral.

Töö tulemusena loodi veebirakenduse prototüüp, mis võimaldab üliõpilaste ja õppejõudude emotsioone tuvastada peaaegu reaalajas. Veebirakenduse loomiseks kasutatakse PostgreSQL andmebaasi ja Streamlit raamistikku.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 32 leheküljel, 5 peatükki, 22 joonist, 3 tabelit.

Abstract

Machine Learning Based Facial Emotion Recognition in Educational Process

The goal of the thesis is to create a method based on machine learning to monitor the emotional state of students and lecturers during the learning process by facial emotions. Four different machine learning methods have been used to create models: SVM, Random Forest, CNN and VGG16 deep convolutional neural network.

Machine learning models were created using Python's Scikit-Learn library, CNN algorithms were realized in Keras running on the TensorFlow backend, all experiments were performed in Google Colab.

To train models, the dataset FER2013 is used in this work. Deepface framework was used to delete mislabeled images and balance the dataset.

Three types of emotion classification are discussed in this master's thesis:

1. Classification into 7 classes: surprise, angry, fear, sad, neutral, happy, disgust
2. Classification into 2 classes: positive (happy) and negative (angry, fear, sad, disgust) emotions.
3. Classification into 2 classes: active (angry, disgust, fear, happy, surprise) and passive (sad) emotions.

The most effective algorithm was VGG16, with the average 10-fold cross-validation accuracy 96% for positive/negative classification model.

As a result of the work, a prototype of a web application has been created, which allows to identify the emotions of students and academic staff in the learning process in near real-time. The PostgreSQL database and Streamlit framework are used to create a web application.

The thesis is in Estonian and contains 32 pages of text, 5 chapters, 22 figures, 3 tables.

Lühendite ja mõistete sõnastik

AdaBoost	<i>Adaptive Boosting</i> , masinõppe algoritm
Box Filter	<i>Box Filter</i> , pilditötluseks kasutatav filter
CERT	<i>Computer Expression Recognition Toolbox</i> , reaalaajas näoilmete tuvastamise süsteem
CNN, ConvNet	<i>Convolutional Neural Network</i> , konvolutsiooniline närvivõrk, mitmekihiliste närvivõrkude klass
FER2013	<i>Facial Expression Recognition 2013</i> , näoemotsioonide andmestik
Gabor	<i>Gabor</i> , pilditötluseks kasutatav filter
GAN	<i>Generative Adversarial Network</i> , generatiivne võistlusvõrk
GentleBoost	<i>GentleBoost</i> , masinõppe algoritm
HBCU	<i>Historically Black College/University</i> , ajalooliselt <i>black</i> kolledž/ülikool Ameerikas
KNN, kNN	<i>K-Nearest Neighbors</i> , k-lähima naabri masinõppe algoritm
RandomForest (RF)	<i>Random Forest</i> , juhuslik mets, masinõppe algoritm
ReLU	<i>Rectified Linear Unit</i> , mittenegatiivne lineaarfunktsioon
SVM	<i>Support Vector Machine</i> , tugivektormasin, masinõppe algoritm
VGG16	<i>Very Deep Convolutional Networks for Large-Scale Image Recognition</i> , 16-kihiline sügav konvolutsiooniline närvivõrk

Sisukord

1 Sissejuhatus	9
1.1 Probleemi taust	9
1.2 Eesmärk ja ülesanded	11
1.3 Ülevaade tööst	12
2 Metoodika.....	13
2.1 Andmete kirjeldus.....	13
2.2 Näoemotsioonide klassifitseerimine.....	16
2.3 Näoemotsioonide tuvastamise protsessi kirjeldus	17
2.4 Algoritmide kirjeldus.....	18
2.4.1 Klassikalised algoritmid	19
2.4.2 Sügavõppe algoritmid.....	20
2.5 Klassifitseerimise moodsused	23
2.6 Tööriistad ja vahendid	25
3 Töö tulemused	27
3.1 Andmete saamine ja eeltöötlus	27
3.2 Eksperimendid	28
3.3 Eksperimentide tulemuste analüüs	34
3.3 Näoemotsioonide tuvastav rakendus	35
4 Tulemuste analüüs ja järeldused.....	37
5 Kokkuvõte	39
Kasutatud kirjandus	41
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	43

Jooniste loetelu

Joonis 1. Emotsioonide jaotus 2D-ruumis Russelli ja Schereri uuringu põhjal. [5].....	10
Joonis 2. Piltide arv emotsioonitüüpide kaupa treeningandmetes.	13
Joonis 3. Piltide arv emotsioonitüüpide kaupa testandmedetes.....	14
Joonis 4. Näide "halbade" piltidest.	14
Joonis 5. Treeningkomplektist valitud fotodel emotsioonide tuvastustäpsuse statistika.	15
Joonis 6. Testimiskomplektist valitud fotodel emotsioonide tuvastustäpsuse statistika.	15
Joonis 7. Andmed csv failist (<i>DataFrame</i>).	15
Joonis 8. Emotsioonide kaks sõltumatut dimensiooni: valents (negatiivne-positiivne) ja aktiivsus (madal-kõrge). [14]	17
Joonis 9. Konvolutsiooni operatsioon. [17].....	21
Joonis 10. Funktsioon ReLU. [17]	21
Joonis 11. Konvolutsioonilise närvivõrgu näide. [17].....	22
Joonis 12. VGG16 arhitektuur. [18]	23
Joonis 13. Veamaatriks (<i>confusion matrix</i>). [19]	24
Joonis 14. Andmestiku näide: csv andmed on salvestatud <i>Dataframe</i> -isse	27
Joonis 15. Näide andmete teisendamisel SVM ja RandomForest algoritmidele	27
Joonis 16. Näide CNN-i mudeli loomise protsessist 7 emotsiooni juhul.	32
Joonis 17. Näide CNN-i mudeli loomise protsessist 7 emotsiooni juhul– veamaatriks ja klassifitseerimismõõdikud.	33
Joonis 18. Näide funktsiooni <code>cross_validate</code> kasutamisest aktiivse/passiivse klassifitseerimise SVM-mudeli juhul.	33
Joonis 19. VGG16 arhitektuuri näide 7 emotsiooni juhu.	34
Joonis 20. Emotsioonituvastuse rakenduse prototüüp.	35
Joonis 21. ERD diagramm.	36
Joonis 22. Tabel <code>teacher_emotion</code>	36

Tabelite loetelu

Tabel 1. SVM ja RF hüperparameetrid kolme tüüpi emotsioonide klassifitseerimiseks	30
Tabel 2. CNN hüperparameetrid kolme tüüpi emotsioonide klassifitseerimiseks	31
Tabel 3. Emotsioonide klassifitseerimise mudelite täpsuste võrdlus	34

1 Sissejuhatus

Tänapäevases infotehnoloogiat kasutavas maailmas peetakse üha enam videokonverentse, koosolekuid, konsultatsioone, loenguid, kasutades selliseid programme nagu *Zoom*, *Microsoft Teams*. Õpetamise taset ja õppijate huvi videokonverentsil või klassiruumis saab hinnata mitte ainult kuulajate kaasatuse taseme, vaid ka nende emotsionaalse tausta kaudu.

Emotsionaalselt positiivse õppimise tausta all mõistetakse tingimusi, mil õppijal on soov suhelda kolleegide ja õppejõuga, ta on valmis aktiivselt töötama, ei tunne ebamugavust ja kohmetust, ta on kindel oma võimetes, tema enesehinnang muutub adekvaatseks. Avatus uute asjade vastuvõtmisele, huvi õppetöö vastu, soov ennast näidata äratab loomingulist potentsiaali. Sellistes töötingimustes teadvustab ja mõistab üliõpilane õppimise eesmärgid ning tal kujuneb vajadus isikliku ja tööalase kasvu järele.

Emotsionaalselt positiivne taust suurendab üliõpilaste töövõimekust, õppimise efektiivsust ja taset, üliõpilaste õppe edukust ja materjali omandamise kvaliteeti ning, mis kõige tähtsam, õppimistahet.

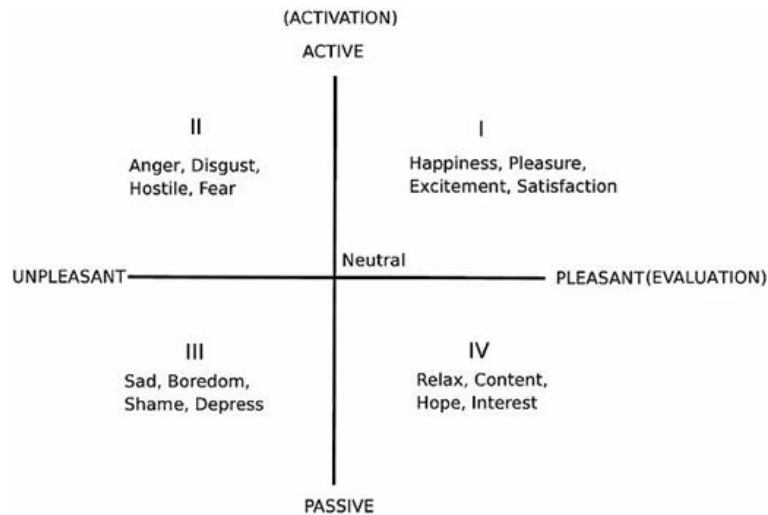
1.1 Probleemi taust

Nägu on inimkeha kõige ilmekam ja peamine interaktiivne osa. See võib sõnadeta edasi anda palju emotsioone.

Emotsioonide uurija Paul Ekmani sõnul [1] on mitmeid emotsioone, mis kvalitatiivselt erinevad neid tekitavate sündmuste, hinnangu, käitumisreaktsioonide ja füüsilise kogemuse poolest. Ta tõestas, et maailma erinevad rahvad, olenemata rassist, keelest ja kultuurilisest arengust, tunnevad ära vähemalt kuus põhiemotsiooni (viha, hirm, vastikus, kurbus, üllatus ja rõõm), kuigi nende iseloomus võib esineda kultuurilisi erinevusi.

Ekman [2] pakkus välja ka meeldiv-ebameeldiv ja aktiivne-passiivne skaalad, millest piisab emotsioonide erinevuse tabamiseks. Russell [3] töötas välja interpersonaalse ringmudeli (*Circumplex model*) ja oletas, et kõiki emotsioone saab paigutada ringi, mida

juhivad kaks sõltumatut dimensiooni: valents (meeldiv-ebameeldiv) ja aktiivsus (aktiivne-passiivne). Scherer [4] on Russelli mudelit värskendanud, et esindada rohkem erinevaid emotsioone. Russelli ja Schereri uuringu põhjal võtsid Hack ja Jackson [5] kokku emotsioonide jaotuse kahes dimensioonis, nagu on näidatud Joonisel 1.



Joonis 1. Emotsioonide jaotus 2D-ruumis Russelli ja Schereri uuringu põhjal. [5]

Näoilme äratundmine haridusprotsessis on muutunud kuumaks uurimisvaldkonnaks. Paljudes artiklites käsitletakse klassikalise masinõppe ja süvaõppe lähenemisviise. Liu ja Wang [6] rakendasid AdaBoost-SVM algoritmi emotsioonide tuvastamiseks e-õppe keskkonnas. Nad saavutasid 95% tuvastustäpsuse PIE (*Pose, Illumination, Expressions*) nägude andmebaasil.

Whitehill jt [7] pakkusid välja lähenemist, mis tuvastab õppijate kaasamise nende näoilmete järgi, võrdlesid õppijate kaasamise automaatset hinnangut ning eeltesti ja järeltesti sooritamise tulemuslikkust, et hinnata kaasamise tuvastamissüsteemi tõhusust. Keskmise täpsuse, mis saavutati *Box Filter* rakendamisel Gentle-Boost algoritmi abil, *Gabor* filtri rakendamisel SVM algoritmi abil, CERT (*Computer Expression Recognition Toolbox*) väljundandmetel multinomiaalse logistilise regressiooni abil, oli vastavalt 0.714, 0.728 ja 0.729 HBCU andmestikul.

Ayvaz jt [8] töötasid välja näoemotsioonide tuvastamissüsteemi (FERS, *Facial Emotion Recognition System*), et tuvastada emotsionaalseid seisundeid ja motiveerida üliõpilasi e-õppes. Süsteemis kasutatakse 4 masinõppe algoritmi (SVM, KNN, juhuslik mets ehk Random Forest, klassifikatsiooni- ja regressioonipuud) ning parimad täpsusskoorid 98,24% saadi SVM algoritmi kasutades.

Lasri jt [9] esitlesid näoilmete tuvastamissüsteemi, mis aitab õpetajal ära tunda õpilaste arusaama oma esitlusest. Konvolutsioonilise närvivõrgu mudeli abil saavutasid nad FER2013 andmestikul 70% täpsuse.

Sathik jt [10] tõestasid, et üliõpilaste näoilmed on olulises korrelatsioonis nende emotsioonidega, mis võib aidata õppejõududel otsustada üliõpilaste kaasatust ja mõistmist.

Seotud tööd kokku võttes võib järeldada, et ennustavate mudelite loomiseks kasutati väga erinevaid andmekogumeid ja masinõppe algoritme. Selles töös kasutatakse andmestikku FER2013 ning luuakse rakendus tudengite ja õppejõudude näoemotsioonide tuvastamiseks õppeprotsessi käigus. Rakendus liigitab näoilmeid kahel viisil. Esimene viis hõlmab seitsme väljendi klassifitseerimist: viha, rõõm, kurbus, üllatus, hirm, neutraalne, vastikus. Teine liigendamiseviis on aktiveerimise-valentsi skaalal: positiivne versus negatiivne, aktiivne versus passiivne [11].

1.2 Eesmärk ja ülesanded

Lõputöö eesmärk on luua masinõppel põhinev meetod üliõpilaste ja õppejõudude emotsionaalse seisundi jälgimiseks õppeprotsessi käigus näo emotsioonide järgi.

Ülesanded:

- Läbi viia viimase kolme aasta jooksul avaldatud näo emotsioonide automaatse tuvastamisega ja selle meetodi õppeprotsessis rakendamise seotud publikatsioonide analüüs.
- Teha erinevate näo emotsioonide tuvastamise algoritmide ja meetodite võrdlusanalüüs.
- Luua näo emotsioonide põhinedes võimalikult efektiivne mudel üliõpilaste ja õppejõudude emotsionaalse seisundi määramiseks õppeprotsessi käigus.
- Luua veebirakendus üliõpilaste ja õppejõudude emotsionaalse seisundi jälgimiseks õppeprotsessi ajal.

1.3 Ülevaade tööst

Mudelite loomiseks ja testimiseks kasutatakse andmestikku *Facial Expression Recognition 2013 (FER2013)*, mis sisaldab 35 887 pilti seitse emotsioonitüüpi (viha, rõõm, kurbus, üllatus, vastikus ja hirm pluss üks neutraalne) kohta.

Mudelite koostamiseks kasutatakse klassikalisi ja sügavõppe klassifitseerimise algoritme nagu SVM, Random Forest ja närvivõrke. Masinõppe mudelid loodi Pythoni Scikit-Learn teegi abil, CNN algoritmid realiseeriti TensorFlow tagarakendusel töötavas Kerases.

Töö tulemusena luuakse üliõpilaste ja õppejõu kontaktõppe protsessis emotsionaalse seisundi jälgiv rakendus, mille abil on võimalik:

1. määrata üliõpilaste ja õppejõu emotsionaalne seisund õppeprotsessi käigus;
2. määrata kontaktõppe keskkonna üldine emotsionaalne taust;
3. fikseerida võimalikud emotsionaalse seisundi häired (näiteks, hirm, paanika, rahutus, mure).

Veebirakenduse loomiseks kasutatakse PostgreSQL andmebaasi ja Streamlit raamistikku. Õppejõud ja üliõpilane autentivad ja autoriseerivad ennast süsteemi sisselogimisel, valivad õppeaine ja vajutavad nuppu Start. Emotsioonid tuvastatakse iga 15 sekundi järel ja salvestatakse andmebaasi. Rakenduses realiseeritakse samuti saadud andmete visualiseerimist, mida õppejõud saab vaadata, kasutades erinevaid andmete valimise kriteeriume.

Magistritöö tulemusena loodud rakendus annab õppejõududele tagasiside tudengite õppimisele pühendumise kohta, mis võimaldab õpetamisprotsessi korrigeerida ja õppekvaliteedi tõsta.

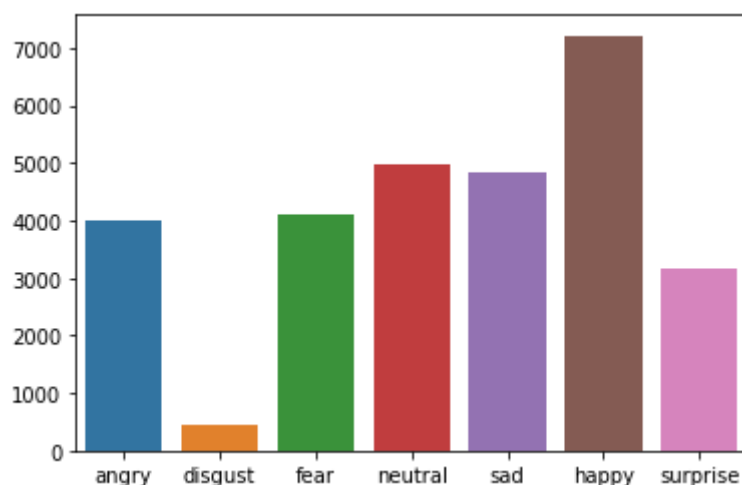
2 Metoodika

2.1 Andmete kirjeldus

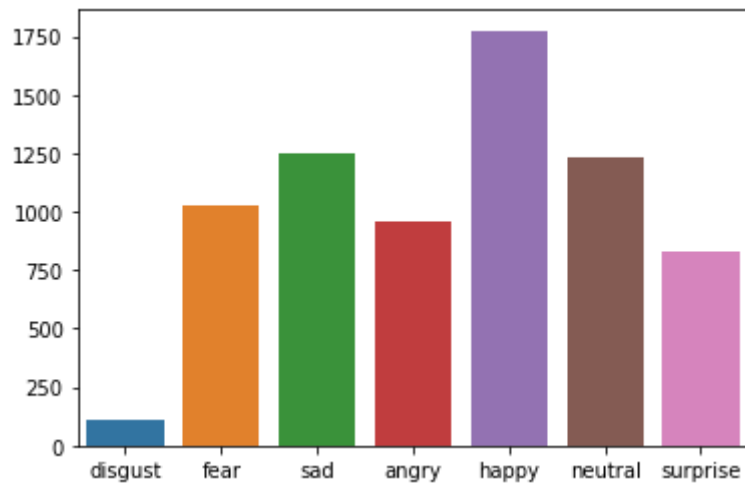
Mudelite treenimiseks antud töös kasutatakse andmestikku FER2013 [12] mida esitleti rahvusvahelisel masinõppe konverentsil (*International Conference on Machine Learning ICML*) 2013. aastal ja mis kujunes etaloniks suuremale arendajate kogukonnale emotsioonituvastuse valdkonnas.

See andmestik sisaldab 35 887 pilti resolutsiooniga 48×48 pikslit, millest enamik on tehtud päriseluga sarnastes tingimustes. Iga pilt on liigitatud üheks seitsmest emotsioonitüübist: üllatus (*surprise*), hirm (*fear*), õnn (*happy*), viha (*angry*), vastikus (*disgust*), kurbus (*sad*) ja neutraalne seisund (*neutral*).

Andmed on juba eelnevalt jagatud treening- (sisaldab 28709 pilti) ja testandmeteks (7718 pilti). Joonistel 2 ja 3 on esitatud piltide jaotus emotsioonitüüpide kaupa treening-ja test andmetes vastavalt.



Joonis 2. Piltide arv emotsioonitüüpide kaupa treeningandmetes.



Joonis 3. Piltide arv emotsioonitüüpide kaupa testandmedetes.

Selle andmestikuga töötamisel ilmnud järgmised probleemid:

- Suur erinevus piltide arvus erinevates kategooriates. Näiteks õnnelike emotsioonidega seotud kujutisi on 7215 treening- ja 1774 testandmetes, kuid vastikuse emotsioone on ainult 436 treening- ja 111 testandmetes vastavalt.
- Mõne pildi kvaliteet on madal. FER2013 sisaldab näota fotosid, uniseid nägusid, teksti pilte, valesti märgistatud pilte (Joonis 4).



Joonis 4. Näide "halbade" piltidest.

Andmete eeltöötlemise etapis esialgu kontrolliti kõiki fotosid face_recognition teegi abil ja kõik mitte-näopildid eemaldati andmekogumist.

Valesti märgendatud piltide kustutamiseks ja andmekogumi tasakaalustamiseks kasutati Deepface'i raamistikku [13]. Kasutades seda raamistikku, valiti emotsioonituvastuse seisukohast 1300 parimat pilti treeningkomplekti igast emotsioonirühmast välja ja kõik pildid testimiskomplektist.

Joonisel 5 on näidatud treeningandmete komplekti igast emotsioonirühmast 1300 valitud pildi järgi leitud minimaalne, maksimaalne ja keskmine emotsiooni tuvastustäpsus. Vastikuse (*disgust*) emotsiooni korral jäeti kõik pildid alles, kuna neid on alla 1300, minimaalne tuvastustäpsus selles rühmas on 0,000287.

Statistic of train dataframe				
	result			
	min	max	mean	count
emotion				
angry	91.345841	100.0	98.120776	1300
disgust	0.000287	100.0	83.539994	340
fear	90.716803	100.0	97.730988	1300
happy	99.995530	100.0	99.999242	1300
neutral	98.523354	100.0	99.615305	1300
sad	83.267819	100.0	95.152093	1300
surprise	97.575790	100.0	99.555911	1300

Joonis 5. Treeningkomplektist valitud fotodel emotsioonide tuvastustäpsuse statistika.

Joonisel 6 on näidatud minimaalne, maksimaalne ja keskmine tuvastamise täpsus igas emotsioonirühmas testkogumist valitud piltide puhul.

Statistic of test dataframe				
	result			
	min	max	mean	count
emotion				
angry	4.688283e-11	100.000000	43.816124	647
disgust	1.677907e-10	100.000000	47.812765	86
fear	3.964922e-11	100.000000	40.749175	645
happy	9.304527e-12	100.000000	78.520579	1381
neutral	3.212561e-08	99.999917	55.485641	929
sad	3.356156e-10	99.999535	35.303039	667
surprise	9.349889e-10	100.000000	71.488339	603

Joonis 6. Testimiskomplektist valitud fotodel emotsioonide tuvastustäpsuse statistika.

Tulemusena valiti FER2013 komplektist kokku 8140 pilti treeningandmetest ja 4577 pilti testimisandmetest ning ühendati treening- ja testimisandmed üheks andmekogumiks.

Järgmiseks muudeti Kerase abil nägude suurus 24x24 piksliteks ja teisendati numbrite massiiviks pikkusega $576 = 24 \cdot 24$.

Andmete eeltöötuse tulemusena loodi 4 csv andmefaili, mille struktuur on esitatud Joonisel 7

	emotion	pixels	Usage	label	positive_negative	active_passive
12203	surprise	68 121 129 129 118 121 133 160 158 151 140 139...	Testing	6		active
7769	surprise	139 169 178 179 177 169 168 172 171 169 169 16...	Training	6		active
996	angry	61 101 117 129 138 153 163 168 172 177 182 184...	Training	0	negative	active
6138	sad	14 6 25 76 74 105 124 136 144 146 141 141 154 ...	Training	5	negative	passive
7392	surprise	154 149 142 138 123 63 99 100 56 55 82 99 87 9...	Training	6		active
11262	neutral	84 85 99 116 138 155 171 183 185 186 189 194 1...	Testing	4		
12520	surprise	201 185 184 182 185 186 190 188 179 176 178 17...	Testing	6		active
8022	surprise	51 37 40 47 64 74 65 88 109 152 165 168 157 16...	Training	6		active
8418	angry	36 25 21 60 135 172 191 199 210 214 213 213 20...	Testing	0	negative	active
4785	neutral	38 51 62 80 96 111 123 133 143 157 176 182 180...	Training	4		

Joonis 7. Andmed csv failist (*DataFrame*).

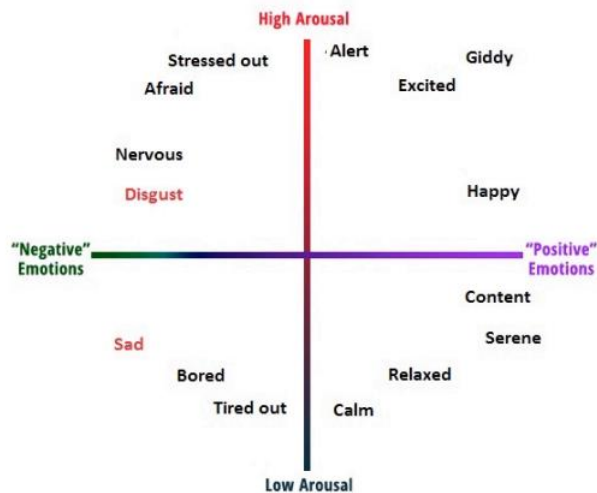
1. Peamine andmestikufail piltidega suurusega 24x24 pikslit mida töös kasutati kõigi algoritmide jaoks: treeningandmete igast kategooriast valitud 1300 parima tuvastustäpsusega nägudega pilti, välja arvatud *disgust* kategooria (340 pilti) ja testimisandmetest valitud kõik nägudega pildid, välja arvatud *happy* kategooria, kust valiti 1000 pilti, et emotsioonitüüpide järgi andmeid paremini tasakaalustada. Andmestiku maht: 8140 treeningandmetest pilti ja 4577 testimisandmetest pilti.
2. Peamise andmestikufaili teisendatud variant piltidega suurusega 48x48 pikslit parimate mudelite ja mudeli VGG16 treenimiseks ja testimiseks. Andmestiku maht: 8140 treeningandmetest pilti ja 4577 testimisandmetest pilti.
3. Andmestikufailid piltidega suurustega 1) 24x24 pikslit ja 2) 48x48 pikslit: *face_recognition* teegi abil tuvastatud nägudega pildid parimate mudelite ja mudeli VGG16 treenimiseks ja testimiseks. Andmestiku maht: 20001 treeningandmetest pilti ja 4958 testimisandmetest pilti.

2.2 Näoemotsioonide klassifitseerimine

Käesolevas magistritöös on käsitletud 3 tüüpi emotsioonide klassifitseerimist:

1. Klassifitseerimine 7 emotsiooni klassiks: üllatus, viha, hirm, kurbus, neutraalne, rõõm, vastikus (*surprise, angry, fear, sad, neutral, happy, disgust*).
2. Klassifitseerimine 2 emotsiooni klassiks: positiivsed (rõõm) ja negatiivsed (viha, hirm, kurbus, vastikus) emotsioonid.
3. Klassifitseerimine 2 emotsiooni klassiks: aktiivsed (viha, vastikus, hirm, rõõm, üllatus) ja passiivsed (kurbus) emotsioonid.

Emotsioonide positiivseks/negatiivseks klassideks jaotamist määrati Joonisel 8 toodud emotsioonide graafiku põhjal.



Joonis 8. Emotsioonide kaks sõltumatut dimensiooni: valents (negatiivne-positiivne) ja aktiivsus (madal-kõrge). [14]

Emotsioonide aktiivseks/passiivseks klassideks jaotamiseks kasutati emotsioonide jaotust kahemõõtmelises ruumis Russelli ja Schereri uuringu põhjal [5]. Hindamise skaala mõõdab inimese tunnet meeldivast ebameeldivani, samas kui aktivatsiooni skaala mõõdab, kui tõenäoline on, et inimene kavatseb tegutseda emotsionaalses seisundis aktiivsest passiivseni. Kahe dimensiooni ristumiskoht vastab neutraalsele seisundile (vt Joonis 1 eespool sissejuhatuses).

2.3 Näoemotsioonide tuvastamise protsessi kirjeldus

Emotsioonide määramisel tuvastavad algoritmid esmalt fotodel või videotes olevad näod ja seejärel tuvastavad mikroväljendused, analüüsides näo punktide vahelisi seoseid. Emotsioonituvastusalgoritme on palju, näiteks *Viola-Jonesi* meetod, mis tugineb skaneerimisakna põhimõttel, kus kasutatakse ristkülikukujulisi tunnuseid, mida nimetatakse *Haari* primitiivideks [15]. Käesolevas töös loodud veebirakendus kasutab seda algoritmi näo tuvastamiseks videovoos.

Inimese emotsioonide avaldumine väljendub näo miimiliste väljenduste kaudu, mis ilmnevad näo võtmepunktide muutumisel. Nende punktide asetus ja määratlemine on emotsioonide määramisel peamine ülesanne. Praegu on selles valdkonnas palju arendusi, kus arendajad otsivad optimaalseid lahendusi emotsioonituvastuse täpsuse parandamiseks.

Olemasolevatel lahendustel on mitmeid puudusi, näiteks paljud algoritmid ja mudelid ei ole vaatamiseks saadaval (peidetud), puudub võimalus lahendust muuta, täiendada, puudub võimalus mudeli uutel andmetel üle treenimiseks.

Käesolevas töös emotsioonide tuvastamine toimub mitmes etapis: pildi ettevalmistamine (etapid 1.-4.), mudeli ettevalmistamine (5. etapp), mudeli rakendamine klassifitseerimisülesande lahendamiseks (6. etapp):

1. Näo asukoha pildil määratlemine.
2. Pildi skaleerimine suurusele 24x24 või 48x48 pikslit. Töös rakendati kõik algoritmid kõigepealt andmestikule piltide suurusega 24x24 pikslit.
3. Värvilise pildi teisendamine halliskaalale.
4. Näopildi normaliseerimine (iga piksli väärtus jagatakse 255-ga), mis on vajalik pildi ettevalmistamiseks närvivõrgus töötlemiseks ja tausta eemaldamiseks.
5. Mudeli treenimine. Töös käsitletakse nii klassikalisi mudeleid (SVM, Random Forest) kui ka närvivõrke (CNN).
6. Optimaalse mudeli rakendamine. CNN näitas eksperimentides parimat emotsioonide tuvastamise tulemust kõigi kolme klassifikatsiooni tüübi järgi.

2.4 Algoritmide kirjeldus

Klassifitseerimine on juhendatud masinõppe ülesanne. On olemas treeningandmestik, kus objektid esitatakse nende tunnuste kirjelduse (pilt - numbrite vektor) ja klassi märgendite (emotsioonide tüüp) kujul. Ülesanne seisneb sellise algoritmi leidmises, mis määrab iga uue pildiobjekti jaoks kõrgeima täpsusega selle objekti klassi märgendi.

Populaarsed klassifitseerimisalgoritmid:

- Logistiline regressioon
- Tugivektormasin (*Support vector machine*, SVM)
- Naiivne Bayes
- Otsustuspuud
- Ansambli mudelid (Random Forests, Gradient Boosted Trees)
- KNN
- Närvivõrgud
- jt

2.4.1 Klassikalised algoritmid

Selles töös käsitleb autor selliseid klassikalisi algoritme nagu SVM ja Random Forest. Neid algoritme vaadeldi, sest töös kasutatakse lisaks mitmeklassilisele klassifitseerimisele ka binaarset emotsioonide klassifitseerimist (positiivne/negatiivne, aktiivne/passiivne).

SVM

Tugivektormasin või SVM (ingl. *Support Vector Machine*) meetod on lineaarne algoritm, mida kasutatakse klassifitseerimise ja regressiooni ülesannetes. See algoritm on praktikas laialdaselt kasutatav ja suudab lahendada nii lineaarseid kui ka mittelineaarseid ülesandeid. SVM-i töö sisu: algoritm loob joone või hüpertasandi, mis jagab andmed klassidesse.

SVM ei toeta multiklassilist klassifitseerimist. See toetab binaarset klassifitseerimist ehk andmete jagamist kahte klassi. Multiklassilise klassifitseerimise puhul kasutatakse multiklassifikaatori probleemi jagamise põhimõtet binaarklassifikaatori mitmeks ülesandeks.

Idee on kujutada andmepunktid mitmemõõtmelisse ruumi, et saada iga kahe klassi vahel vastastikune lineaarne eraldus. Seda nimetatakse *One-to-One* (üks-ühele) lähenemisviisiks, mis jagab multiklassi probleemi mitmeks binaarse klassifitseerimise probleemiks iga klassipaari jaoks. [16]

Teine lähenemisviis, mida saab kasutada, on *One-to-Rest*. Selle lähenemisviisi puhul määratakse binaarne klassifikaator iga klassi jaoks.

1. *One-to-Rest* lähenemisviisi korral võib klassifikaator kasutada m SVM-i. Iga SVM ennustab kuulumist ühte m klassist.
2. *One-to-One* lähenemisviisi puhul võib klassifikaator kasutada $m*(m-1)/2$ SVM-i. [16]

Random Forest

Juhuslik mets (*Random Forest*, RF) on juhendatud masinõppe algoritm. Seda saab kasutada nii klassifitseerimise kui ka regressiooni ülesannete lahendamiseks. RF loob

juhuslikult valitud andmestiku jaoks otsustuspuid, saab igalt puult prognoosi ja valib hääletamise teel parima lahenduse.

Random Forest on meetod (mis põhineb "jaga ja valitse" lähenemisel), mis kasutab juhuslikult jagatud andmestikul loodud otsustuspuude ansamblit. Selliste puuklassifikaatorite komplekt moodustab metsa. Klassifitseerimisülesandes hääletab iga puu ja lõpptulemusena valitakse populaarseim klass.

2.4.2 Sügavõppe algoritmid

Sügavõppe on tehisnärvivõrkudel põhinev masinõppe tüüp. Õppeprotsessi nimetatakse sügavaks, kuna tehisnärvivõrkude struktuur koosneb mitmest sisend-, väljund- ja peidetud kihist.

Erinevaid närvivõrkude tüüpe on palju, näiteks rekurrentne närvivõrk (RNN - *Recurrent Neural Networks*), konvolutsiooniline närvivõrk (CNN - *Convolutional Neural Networks*), mitmekihilised täisühendusega närvivõrgud (*Fully-Connected Neural Networks*, FCNN).

Funktsionaalselt jagatakse närvivõrgud kihtideks:

- Sisendkiht saab andmekogumi;
- Peidetud kihid teostavad sisendparameetrite põhjal arvutusi. Kõige sagedamini sisaldavad närvivõrgud mitut peidetud kihti;
- Väljundkiht väljastab arvutuse tulemuse.

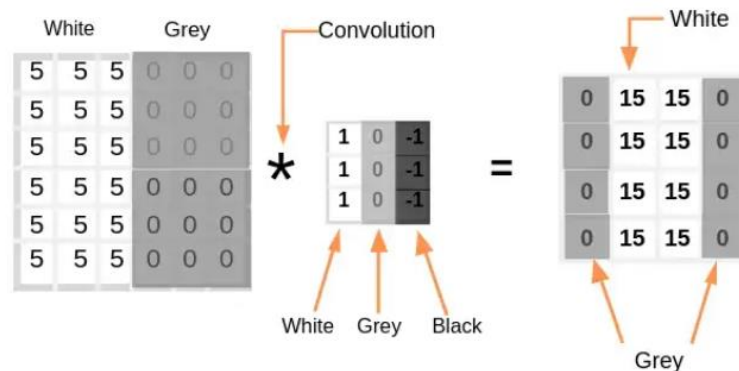
Konvolutsiooniline närvivõrk (CNN) on peamine tööriist objektide, fotodel olevate nägude ja kõnetuvastuse klassifitseerimiseks ja äratundmiseks.

Piltidel on kõrgus ja laius, seega need esitatakse kahemõõtmelise struktuuri (maatriksi) abil. Kui aga kasutada värvilist pilti, siis kasutatakse ka pildi sügavust. Pildi sügavus halli toonides on 1, kui see oleks RGB (*red green blue*), siis sügavus on 3. [17]

CNNi põhimõisted:

- Kernel (tuum) on kaalumaatriks, mis liigub läbi lähtepildi ja leiab objektide teatud tunnused (jooned, servad, tekstuurid jne). Kõige sagedamini määratakse tuuma suurus vahemikus 3x3 kuni 7x7.

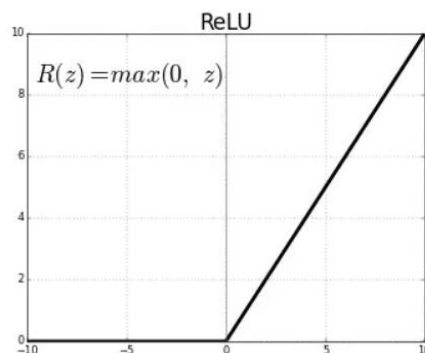
- Filter on kernelite kollektsioon. Filter liigub piki pilti ja määrab, kas teatud otsitav tunnus on selle konkreetses osas olemas.
- Konvolutsiooni operatsioon on filtrielementide ja sisendsignaali maatriksi korrutiste summa (Joonis 9). Kui pildifragmendis on mõni soovitud omadus, annab konvolutsiooni operatsioon väljundis suhteliselt suure väärtusega arvu. Kui omadus puudub, on väljundarv väike. [17]



Joonis 9. Konvolutsiooni operatsioon. [17]

Väljundi kui ka sisendi suuruse säilitamiseks võib kasutada täidistust (*padding*). Täidistus on sisendmaatriksile nullide sümmeetrilise lisamise protsess. Filter võib liikuda piki sisendsignaali maatriksit sammuga, mis erineb ühest. Filtri liigutamise sammu nimetatakse konvolutsiooni sammuks (*stride*). Selleks, et kernelites sisalduvate kaalude treenimine oleks tõhus, tuleb konvolutsiooni tulemustesse sisse viia mõningane nihe (*bias*), mis liidetakse väljundmaatriksi igale elemendile juurde enne aktivatsioonifunktsiooni rakendamist. [17]

Aktivatsioonifunktsiooni kasutamine tagab mudeli mittelinearsuse. Sageli kasutatakse aktivatsioonifunktsioonina ReLU-d (Rectified Linear Unit).

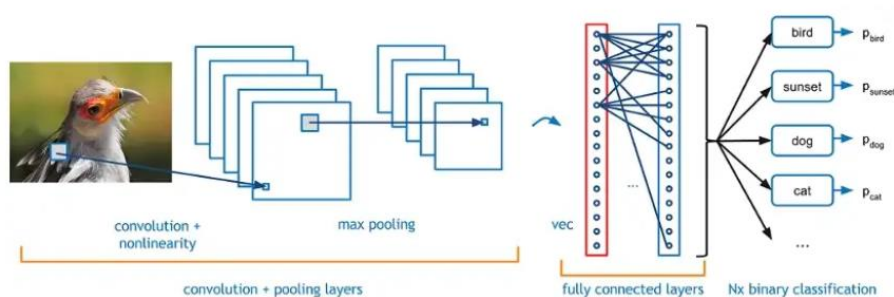


Joonis 10. Funktsioon ReLU. [17]

Joonisel 10 on esitatud funktsioon ReLU, kus sisendväärtused, mis on väiksemad või võrdsed nulliga, muutuvad nulliks ja nullist suuremad väärtused ei muutu.

CNNi kihid (Joonis 11):

- *Input layer* - CNNi sisendkiht peab sisaldama andmeid, näiteks pilte.
- *Convo layer (Convo + ReLU)* - konvolutsioonilist kihti nimetatakse mõnikord tunnuste väljavõtmise kihiks, sest selles kihis võetakse välja pildi tunnused. Convo kiht sisaldab ka ReLU aktivatsiooni, et muuta kõik negatiivsed väärtused nulliga võrdseks.
- *Pooling layer* - ahenduskihti kasutatakse sisendpildi mahu vähendamiseks pärast konvolutsiooni rakendamist. Kasutatakse *max-pooling*-i (maksimaalse väärtuse valimine) ja *average-pooling*-i (maatriksielementide keskmine väärtus) operatsioone. Seda kihti kasutatakse kahe konvolutsioonilise kihi vahel.
- *Fully Connected (FC) layer* - täielikult ühendatud kiht sisaldab kaalusid, nihkeid ja neuroneid. See ühendab ühe kihi neuroneid teise kihi neuronitega. Seda kasutatakse piltide klassifitseerimiseks erinevate kategooriate vahel õppimise teel.
- *Softmax/Logistic layer* - *Softmax* või logistiline kiht on CNN-i viimane kiht. Logistilist kihti kasutatakse binaarse klassifikatsiooni jaoks ja softmaxit mitmeklassilise klassifikatsiooni jaoks.
- *Output layer* - määrab objekti klassi. [17]



Joonis 11. Konvolutsioonilise närvivõrgu näide. [17]

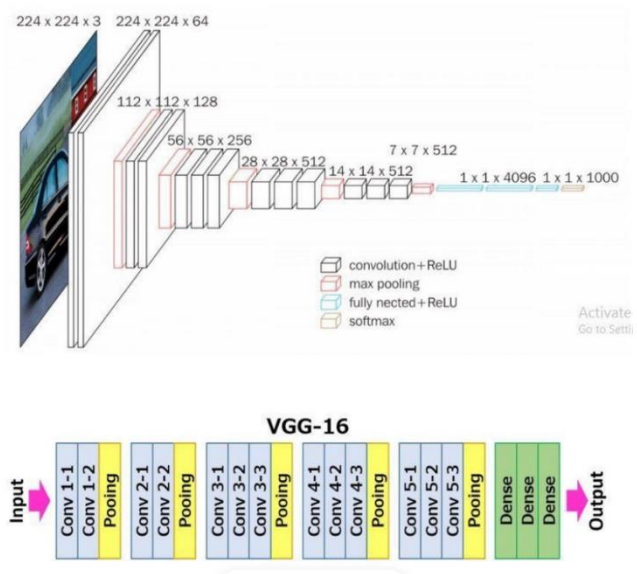
Töös kasutati ka konvolutsioonilist närvivõrku VGG16.

VGG16 on CNN-i ehk konvolutsioonilise närvivõrgu tüüp, mida peetakse tänapäeval üheks parimaks arvutinägemise mudeliks. Selle mudeli loojad (Karen Simonyan &

Andrew Zisserman, Visual Geometry Group, Department of Engineering Science, University of Oxford) hindasid teisi võrke ja suurendasid sügavust, kasutades väga väikeste (3×3) filtritega arhitektuuri. [18]

VGG16 on objektide tuvastamise ja klassifitseerimise algoritm, mis on võimeline klassifitseerima pilte 1000 erinevast kategooriast. [18]

Joonisel 12 on esitatud VGG16 närvivõrgu arhitektuur. Võrk koosneb 16 kihist. Selle närvivõrgu kasutamisel oma klassifitseerimisülesande lahendamiseks saab muuta sisend- ja väljundkihti, mitte kasutada täielikult ühendatud kihte (*Fully Sonnected layers*), lisada oma Convo kihid jne.



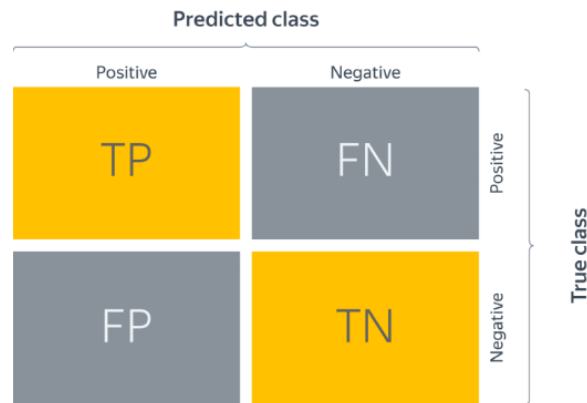
Joonis 12. VGG16 arhitektuur. [18]

2.5 Klassifitseerimise moodsid

Masinõppe ülesannetes kasutatakse mudelite kvaliteedi hindamiseks ja erinevate algoritmide võrdlemiseks moodsid. Klassifitseerimismudelite hindamiseks on palju erinevaid näitajaid. Optimaalne näitaja sõltub probleemist, andmetest ja sellest, mida mudelilt oodatakse.

Moodsidute arvutamisel kasutatakse veamaatriksit (*confusion matrix*) (Joonis 13). Binaarse klassifikatsiooni jaoks kasutatakse järgmisi tähiseid (positiivne/ negatiivne klassidel):

- *true positive* (TP) – õige-positiivsete ennustuste arv;
- *false positive* (FP) – vale-positiivsete ennustuste arv;
- *true negative* (TN) – õige-negatiivsete ennustuste arv;
- *false negative* (FN) – vale-negatiivsete ennustuste arv.



Joonis 13. Veematriks (*confusion matrix*). [19]

Accuracy (õigsus, täpsus) näitab kõigi andmete tõeliste positiivsete ja negatiivsete protsendimäära. Kasutatakse sagedamini, kui andmestik on tasakaalustatud. Selles töös ei ole andmed emotsioonitüüpide järgi täielikult tasakaalus, mistõttu vaadeldakse teisi mõõdikuid.

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

Precision (täpsus) on õigesti ennustatud positiivsete objektide osakaal kõigi positiivse klassiga ennustatud objektide hulgas. Mida vähem valepositiivseid tulemusi mudel võimaldab, seda suurem on selle täpsus.

$$precision = \frac{TP}{TP+FP} \quad (2)$$

Recall (saagis) on õigesti leitud positiivsete objektide osakaal kõigi positiivse klassi objektide hulgas. Mida vähem valenegatiivseid vastuseid, seda suurem on mudeli saagis.

$$recall = \frac{TP}{TP+FN} \quad (3)$$

Mudeleid on väga mugav võrrelda, kui nende kvaliteeti väljendatakse ühe numbriga. Täpsuse-saagise paari puhul on populaarne viis ühendada need üheks mõõdikuks – F1 skoor. F1 skoor arvutab harmoonilise keskmise täpsuse ja täielikkuse vahel ning mõlemad

sõltuvad valepositiivsetest ja valenegatiivsetest tulemustest. F1-hinnangut kasutatakse sagedamini siis, kui andmestik on tasakaalustamata.

$$F1 = \frac{2 * precision * recall}{precision + recall} \quad (4)$$

Mitmeklassilise klassifikatsiooni korral muutub mõõdikute arvutamine keerulisemaks. Iga klassi 1...K jaoks saab arvutada oma veamaatriksi. Seejärel on K veamaatriksitest lõpliku mõõdiku väärtuse saamiseks kaks varianti:

- Mikrokeskmistamine. Arvestatakse veamaatriksi elementide (TP, FP, TN, FN) keskmisi väärtusi binaarklassifikaatorite vahel ja luuakse keskmiste väärtustega veamaatriks. Seejärel arvutatakse täpsus, saagis, F1 skoor vastavalt keskmistatud veamaatriksile.
- Makrokeskmistamine. Arvutatakse täpsus, saagis, F1 skoor iga klassifikaatori jaoks eraldi ja seejärel arvutatakse nende keskmine väärtus. [19]

Töös kasutatakse Accuracy (täpsust) ja F1 skoori.

2.6 Tööriistad ja vahendid

Töös on kasutatud *Google Colaboratory*, *Visual Studio Code*, programmeerimiskeelt Python, Pythoni teke ja raamistikke: *Numpy*, *Pandas*, *OpenCV*, *Scikit-learn (Sklearn)*, *Keras*, *Tensorflow*, *Seaborn*, *Matplotlib*, *Pickle*, *Face_Recognition*, *DeepFace*, *Streamlit*.

*Google Colaboratory*¹ või lihtsalt *Colab*, võimaldab kirjutada ja täita Pythoni koodi brauseris ja annab ligipääsu graafilistele protsessoritele.

NumPy — kasutatakse n-suuruse massiividega töötamiseks. *NumPy* kasutavad sageli teised *TensorFlow*-tüüpi teegid, et teha sisemisi arvutusi tensoritega.

¹ <https://colab.research.google.com/>

Pandas — kasutatakse andmete analüüsimiseks, see on üks populaarsemaid teeke. See pakub palju kasulikke vahendeid andmete kogumiseks, puhastamiseks ja modelleerimiseks.

Scikit-learn on üks populaarsemaid Pythoni masinõppe teeke. Pärast andmete puhastamist ja manipuleerimist *Pandas* või *NumPy* abil kasutatakse masinõppe mudelite loomiseks *Scikit-learn*i.

TensorFlow on üks populaarsemaid Pythoni teeke närvivõrkude loomiseks.

Keras — kasutatakse peamiselt süvaõppe mudelite ja närvivõrkude loomiseks. Kasutab *TensorFlow*'d ja *Theano*t ning võimaldab lihtsasti luua närvivõrke.

Seaborn, *Matplotlib* on Pythoni teegid erinevate visualisatsioonide loomiseks.

OpenCV on pildi- ja videotöötluste teek, mida kasutatakse nende analüüsimiseks.

Face_recognition on teek, mis sisaldab näotuvastusfunktsioone.

Deepface on raamistik näotuvastuseks ja näo atribuutide (vanus, sugu, emotsioonid ja rassiline kuuluvus) analüüsimiseks Pythoni jaoks. See on hübriidne näotuvastussüsteem, mis ühendab kaasaegsed mudelid: *VGG-Face*, *Google FaceNet*, *OpenFace*, *Facebook DeepFace*, *DeepID*, *ArcFace*, *Dlib* ja *SFace*. [20]

Pickle pakub funktsioone ja klasse objektide serialiseerimiseks ja deserialiseerimiseks. Töös kasutatakse seda SVM ja Random Forest mudelite salvestamiseks. CNN-i salvestamiseks kasutatakse JSON- ja h5-andmevorminguid.

Streamlit on Pythoni teek, mis võimaldab luua veebirakendusi näiteks masinõppeks ja andmetöötlusteks.

Visual Studio Code on skriptide kirjutamise redaktor.

Data Augmentation on teisenduste abil lähteandmete valiku laiendamismeetod. Kasutatakse treeningandmete suurendamiseks mudelite loomisel. *TensorFlow*'s on klass *ImageDataGenerator*, mis sisaldab argumente piltide teisendamiseks. Näiteks kasutati selles projektis järgmisi argumente.

`ImageDataGenerator(rescale=1./255, - andmete skaleerimine`

`rotation_range = 10, piltide pööramine (näidatakse kraadides)`

`horizontal_flip = True, pildi peegeldamine horisontaalselt`

`width_shift_range=0.1, nihe laiuses - X-teljel`

`height_shift_range=0.1, nihe kõrgusel - Y-teljel`

Testimisandmetele ja valideerimisandmetele kasutati vaid skaleerimist.

3.2 Eksperimendid

Masinõppe mudelid loodi *Pythoni Scikit-Learn*i teegi abil, CNN algoritmid realiseeriti *TensorFlow* tagarakendusel töötavas *Kerases*, kõik eksperimendid teostati *Google Colabis*.

Enamikul masinõppe ja süvaõppe algoritmidel on mõned häälestatavad parameetrid, mida nimetatakse hüperparameetriteks.

Hüperparameetrite kasutamise abil leiti parim SVM mudel emotsioonide tuvastamiseks.

SVM-i algoritmi jaoks kasutati järgmisi hüperparameetreid:

1. tuuma (ingl. *kernel*) funktsioonid:
 - a. *linear* - sisendandmete skalaarne korrutis
 - b. *rbf* – *radial basis function* – radiaalne tuuma funktsioon. See funktsioon lisab seadistamiseks täiendava hüperparameetri γ (gamma). Kuid erinevalt lineaarse tuuma funktsioonist näitab see funktsioon andmeid kõrgemas dimensioonis ja võib kujutada mittelineaarset eraldamist.
2. C - vigade kontrollimise (regulariseerimise) parameeter.

3. gamma on hüperparameeter, mis määrab otsustuspiiride sujuvust ja kontrollib mudeli varieeruvust. Suur gamma tähendab suuremat otsustuspiiride kõverust, mis viib suure varieeruvuse ja liigse kohandamiseni. Väga väike gamma korral mudel ei suuda andmete "kuju" tabada. Kasutatakse, kui kernel on rbf. [21]

Tavaliselt valitakse C valimiseks väärtused: 0,001, 0,01, 0,1, 1, 10, 100. Gamma jaoks: 0,001, 0,01, 0,1, 1, 10, 100. [22]

Parema Random Forest mudeli leidmiseks kasutati ka hüperparameetreid.

1. n_estimators - puude arv ansamblis
2. max_features - tunnuste arv ühe puu ehitamisel
3. min_samples_leaf - minimaalne objektide arv lehes
4. min_samples_split - minimaalne objektide arv, mis on vajalik sisemise sõlme jagamiseks
5. max_depth - puu maksimaalne sügavus.

Hüperparameetrite valimisel kasutati järgmisi parameetreid ja väärtusi:

SVM:

- kernel: rbf, linear
- C: 0.001, 0.01, 0.1, 1, 10, 100
- gamma: 0.001, 0.01, 0.1, 1, 10, 100

Random Forest:

- n_estimators: 300, 500, 700, 900, 1100, 1300, 1500
- max_depth: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100
- max_features: auto, sqrt, log2
- min_samples_split: 2, 5, 10
- min_samples_leaf: 1

Tabelis 1 on esitatud SVM ja Random Forest hüperparameetritega mudelid, mis näitasid parimaid tulemusi mudelite 3-kordsel ristvalideerimisel (cross-validation) kolme tüüpi emotsioonide klassifitseerimise korral.

Tabel 1. SVM ja RF hüperparameetrid kolme tüüpi emotsioonide klassifitseerimiseks

Tüüp	Algoritm	Valitud hüperparameetrid
7 emotsioonid	SVM	kernel = rbf, C = 100, $\gamma = 0.1$
	RF	n_estimators: 1500, min_samples_split: 2, min_samples_leaf: 1, max_features: auto, max_depth: 20
active / passive	SVM	kernel = rbf, C = 100, $\gamma = 0.01$
	RF	n_estimators: 300, min_samples_split: 2, min_samples_leaf: 2, max_features: auto, max_depth: 80
positive / negative	SVM	kernel = rbf, C = 100, $\gamma = 0.01$
	RF	n_estimators: 300, min_samples_split: 2, min_samples_leaf: 1, max_features: sqrt, max_depth: 80

Konvolutsioonilise närvivõrgu konstrueerimiseks kasutatakse 3 põhikihti: konvolutsiooniline (*convolution*), ahendus- (*pooling*) ja täielikult ühendatud (*fully connected, FC*) kihti. Konvolutsiooni- ja ahenduskihte kasutatakse tunnuste eraldamiseks lähtekujutisest, samas kui täielikult ühendatud kihte kasutatakse kujutise lõplikuks klassifitseerimiseks vastavalt välja võetud tunnustele.

Mudeli koostamiseks kasutati *Keras Tuner*¹ – närvivõrgu hüperparameetrite automaatset optimeerijat.

CNN sisaldab:

- sisendkihti, võrgu sisendkihi suurus on $24 \times 24 \times 1$, vastavalt andmestikust pärit piltide suurusele;
- 3 konvolutsioonilist kihti, mis sisaldavad ahenduskihte, ja üleõpetamise vältimiseks kasutatakse reguleerimismehhanisme, nagu näiteks *Batch Normalization* ja *Dropout*;
- täielikult ühendatud kihti;
- väljundkiht on vektor 7 või 2 elemendist, mis vastavad sisendpildi iga emotsioonitüübi kuulumise tõenäosustele.

¹ https://keras.io/keras_tuner/

Batch Normalization - närvivõrgu peidetud kihtide sisendväärtuste normaliseerimine.

Dropout viskab õppimise ajal närvivõrgust kihtides olevaid neuroneid juhuslikult välja.

Tabelis 2 on esitatud hüperparameetrid, mille määratles *Keras Tuner*.

Tabel 2. CNN hüperparameetrid kolme tüüpi emotsioonide klassifitseerimiseks

Hyperparameter	7 emotsioonid	active/passive	positive/negative
conv_1_filter	48	112	112
conv_1_kernel	3	3	3
conv_2_filter	96	64	128
conv_2_kernel	3	3	3
conv_3_filter	256	224	64
conv_3_kernel	3	6	3
dense_1	240	96	48
learning_rate	0.001	0.001	0.001
decay	1e-06	1e-06	1e-06

Töös kasutatakse CNN-i mudelite loomisel tagasikutset (*callback*) - *EarlyStopping* - mudeli ülesobitamise vältimiseks. *EarlyStoppingul* on erinevad parameetrid/mõõdikud, mida saab muuta, et määrata, millal mudeli treenimine peab lõppema.

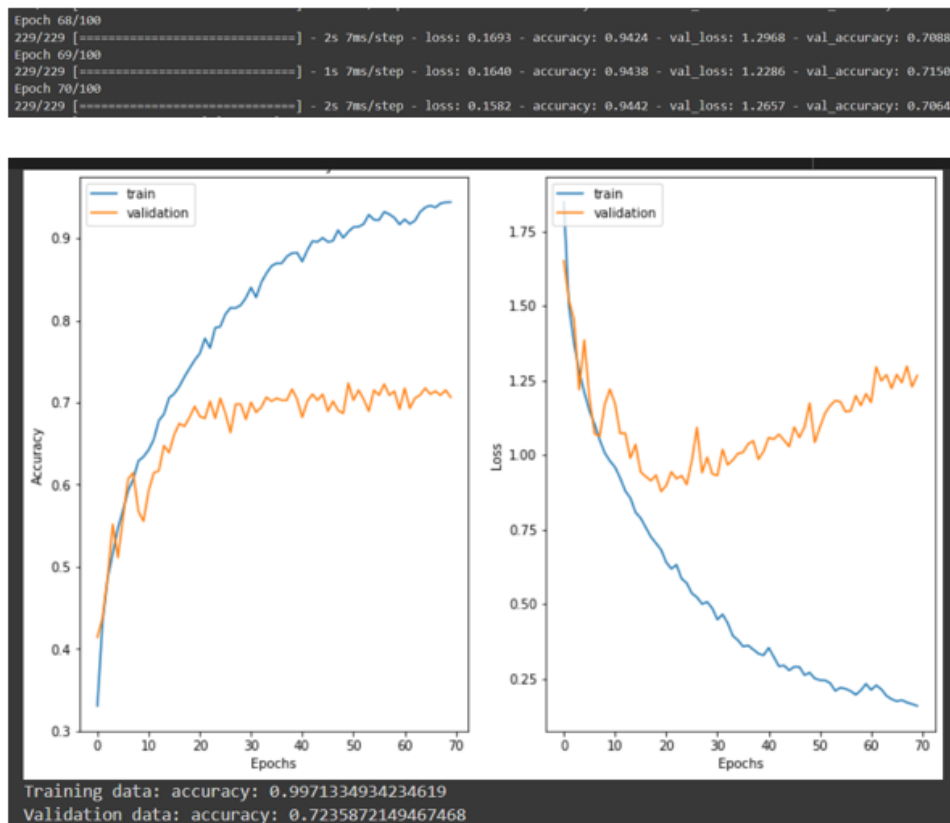
Töös kasutatakse järgmisi mõõdikuid:

```
early_stopping = EarlyStopping(patience=20,  
                                restore_best_weights=True,  
                                monitor='val_accuracy',  
                                min_delta = 0.001,  
                                mode='max')
```

- *monitor*: jälgitav mõõdik.
- *min_delta*: jälgitava mõõdiku minimaalne muutus.
- *patience*: paranemata perioodide arv, mille järel treening peatatakse.
- *mode* - üks režiimidest {"auto", "min", "max"}. *max* - mudeli treenimine peatub, kui jälgitav mõõdik enam ei kasva.
- *restore_best_weights*: kas taastada mudelikaalud jälgitava mõõdiku parima väärtusega epohhist. [23]

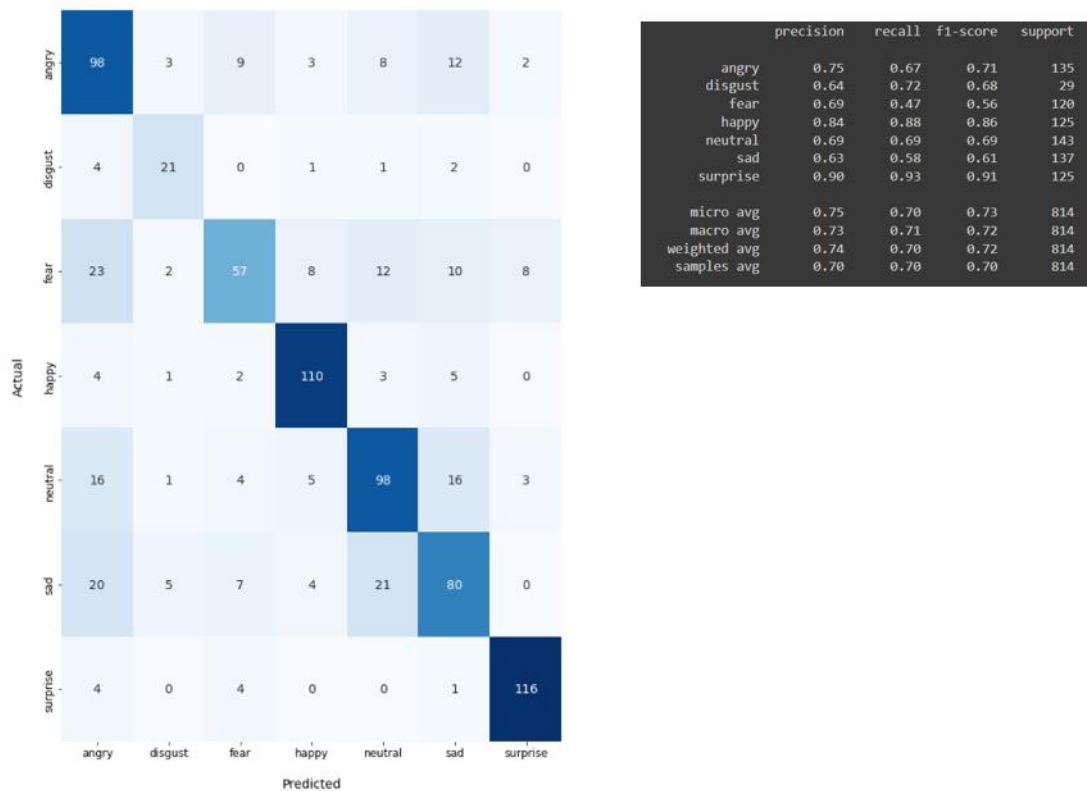
Mudelite ehitamisel kasutati k-kordset ristvalideerimist (*k-fold cross-validation*), millal andmed jagatakse k plokiks, misjärel treenitakse mudelit k-1 plokil ja valideeritakse k-ndal plokil.

CNN-i mudelite loomise igal etapil on võimalik näha iga epohhi mõõdikuid (Joonis 16) ning väljastada vahemudelite täpsuste ja kadude diagramme. Näitajad arvutatakse mudeli treenimise ja valideerimise kohta ning nende seos näitab, kui hästi mudel toimib nende kahe andmekogumi puhul.



Joonis 16. Näide CNN-i mudeli loomise protsessist 7 emotsiooni juhul.

Samuti koostati igal tööetapil veamaatriks ja kuvati aruanne peamiste klassifikatsioonimõõdikute kohta (Joonis 17).



Joonis 17. Näide CNN-i mudeli loomise protsessist 7 emotsiooni juhul- veamaatriks ja klassifitseerimismõõdikud.

SVM-i ja Random Foresti k-kordse valideerimisega mudeli treenimisel kasutati scikit-learn'i teegi funktsioone `cross_val_score`, `cross_validate` ja `cross_val_predict`, et vaadata mõõdikute/näitajate aruannet.

- `cross_val_score` – tagastab ühe mõõdiku, vaikimisi täpsus (*accuracy*)
- `cross_validate` – lubab mudeli hindamiseks määrata mitu indikaatorit (Joonis 18)
- `cross_val_predict` – tagastab näitajad testikomplekti kohta.

	fit_time	score_time	test_accuracy	test_precision	test_recall	test_f1_score
0	13.459497	1.219122	0.839181	0.876271	0.933213	0.903846
1	13.021955	1.080616	0.809942	0.866782	0.904332	0.885159
2	11.956782	1.279538	0.824561	0.876736	0.911552	0.893805
3	13.391535	1.229064	0.817251	0.873043	0.906137	0.889283
4	12.279712	1.425575	0.837719	0.890653	0.911552	0.900981
5	12.399024	1.262039	0.821637	0.871134	0.915162	0.892606
6	12.512793	1.221801	0.821637	0.876307	0.907942	0.891844
7	13.295483	1.007709	0.799708	0.863874	0.893502	0.878438
8	12.877042	1.231059	0.802632	0.865620	0.895307	0.880213
9	13.928763	1.280345	0.839181	0.892226	0.911552	0.901786

Joonis 18. Näide funktsiooni `cross_validate` kasutamisest aktiivse/passiivse klassifitseerimise SVM-mudeli juhul.

Töö käigus viidi läbi ka eksperimendid, kasutades närvivõrku VGG16.

VGG16 närvivõrgu arhitektuur oli optimeeritud järgmisel viisil (Joonis 19): lisati 3 täisfunktsionaalset kihti täiendavate kihtidega BatchNormalization ja Dropout ning väljundkiht, sõltuvalt emotsioonide klassifitseerimise tüübist.

```
def create_model():
    base_model = VGG16(input_shape=(48,48,3), include_top=False)
    model=Sequential()
    model.add(base_model)
    model.add(Dropout(0.5))
    model.add(Flatten())
    model.add(BatchNormalization())
    model.add(Dense(128, kernel_initializer='he_uniform'))
    model.add(BatchNormalization())
    model.add(Activation('relu'))
    model.add(Dropout(0.5))
    model.add(Dense(64, kernel_initializer='he_uniform'))
    model.add(BatchNormalization())
    model.add(Activation('relu'))
    model.add(Dropout(0.5))
    model.add(Dense(32, kernel_initializer='he_uniform'))
    model.add(BatchNormalization())
    model.add(Activation('relu'))
    model.add(Dense(num_classes, activation='softmax'))
```

Joonis 19. VGG16 arhitektuuri näide 7 emotsiooni juhu.

3.3 Eksperimentide tulemuste analüüs

Emotsioonide klassifitseerimise mudelite omaduste hindamiseks kasutati 10-kordset ristvalideerimist (*cross-validation*) ja arvutati täpsuste (*accuracy*) ja F1 skooride keskmised ning segadusmaatriksid (*confusion matrix*). Tulemused on esitatud Tabelis 3.

Tabel 3. Emotsioonide klassifitseerimise mudelite täpsuste võrdlus

	SVM		RF		CNN (24x24)		CNN (48x48)		VGG16 mudel (modifitseeritud)	
	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc
7 emotsioonid	0.60	0.59	0.54	0.57	0.71	0.72	0.69	0.70	0.65	0.67
positive/ negative	0.81	0.87	0.66	0.82	0.91	0.95	0.92	0.94	0.94	0.96
active/ passive	0.69	0.82	0.58	0.83	0.77	0.87	0.79	0.88	0.78	0.88

Tabelist on näha, et kõige efektiivsem oli algoritm VGG16 kõige suurema ristvalideerimise täpsusega 96% positiivse/negatiivse klassifitseerimismudeli korral. Kuid CNN-i mudelid pildisuurusega 24x24 näitasid kolme klassifitseerimise puhul parimaid või parimatele lähedased tulemused.

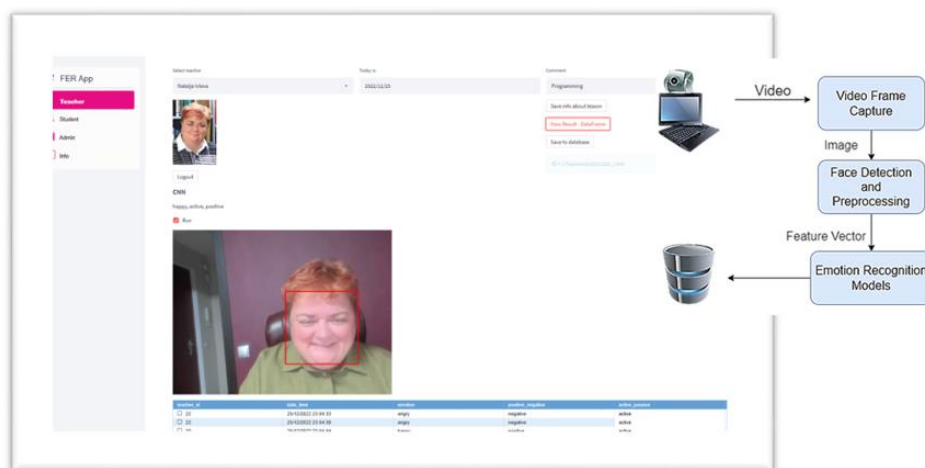
Klassikalisi algoritme (SVM, RF) vaadeldi võrdlemiseks närvivõrkudega (CNN), kuna binaarse klassifikatsiooni jaoks kasutatakse klassikalisi algoritme masinõppes väga sageli, kuid närvivõrgud näitasid emotsioonide tuvastusel suuremat täpsust.

CNN mudelite loomiseks kasutati andmestikku, kus piltide suurus sai 48x48 pikslit. CNN-i abiga loodi uued mudelid, mis näitasid Tabelis 3 esitatud tulemusi.

Kõik kolm parimat CNN-i klassifitseerimismudelit realiseeriti veebirakenduses eesmärgiga määrata üliõpilaste ja õppejõudude emotsionaalne seisund õppimise käigus reaalajas.

3.3 Näoemotsioonide tuvastav rakendus

Töö tulemusena loodi veebirakenduse prototüüp, mis võimaldab näha, kuidas loodud emotsioonide tuvastamise mudelid töötavad (Joonis 20).



Joonis 20. Emotsioonituvastuse rakenduse prototüüp.

Rakenduses on kahte tüüpi kasutajaid: õppejõud ja üliõpilane. Igapähele on loodud sisselogimisvorm. Andmed õppejõudude autentimiseks salvestatakse andmebaasi. Üliõpilane saab õppejõult parooli ja tunnikoodi süsteemi sisenemiseks. Samuti üliõpilane saab oma koodi automaatselt (student_id) ja osaleb emotsioonide määratlemisel anonüümselt. Edaspidi saab realiseerida üliõpilaste andmete salvestamist: perekonnanimi, eesnimi, rühm.

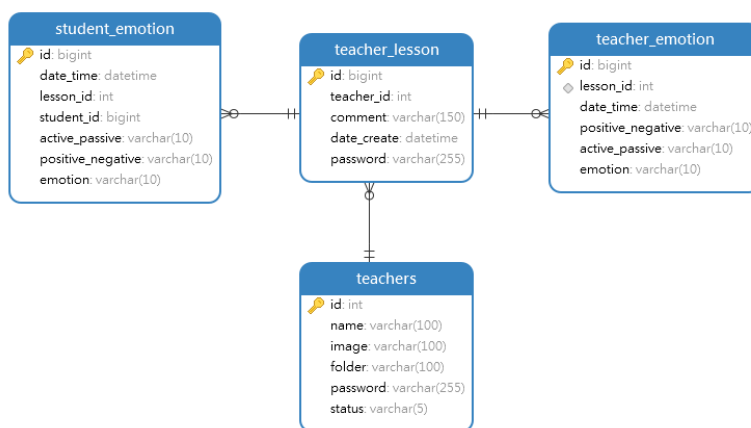
Kasutaja emotsioonituvastussüsteem on realiseeritud veebirakendusena, mis saab videokaamera abil reaalajas pilte. Pildil määratakse inimese nägu ja kui see on leitud,

skaleeritakse see 24x24 piksli suuruseks pildiks ja teisendatakse 256 halli tooniks (hallitoonide värvirežiim, väärtus 0 on must ja väärtus 255 on valge).

Emotsioonide määramiseks mudeli jaoks teisendatakse pilt arvude massiiviks. Mudelid, kasutades massiivi, määravad emotsiooni kolme tüübi järgi: konkreetne emotsioon, positiivne/negatiivne ja passiivne/aktiivne. Saadud andmed salvestatakse andmebaasi.

Andmete salvestamiseks valis autor PostgreSQL relatsiooniandmebaasi, mis asub Virumaa kolledži serveris. Andmebaasis on 4 tabelit.

Joonisel 21 on esitatud andmebaasi ERD diagramm.



Joonis 21. ERD diagramm.

Andmed tabelisse *teachers* lisati *Web Scrapingu* abil Virumaa kolledži veebilehelt (<https://taltech.ee/virumaa-kolledz-kontaktid>). Lehelt Kontaktid võeti välja järgmised andmed: töötaja ees- ja perekonnanimi, foto, staatus - õppejõud või mitte.

Joonisel 22 on näha kuidas andmeid tabelitesse salvestatakse, tabeli *teacher_emotion* näitel. Edaspidi tabelisse lisatakse emotsioonide tuvastustõenäosustega veerg.

id [PK] bigint	lesson_id integer	date_time timestamp without time zone	positive_negative character varying (10)	active_passive character varying (10)	emotion character varying (10)
66	4	2022-12-25 21:58:47	negative	active	angry
67	4	2022-12-25 21:59:00	negative	active	angry
68	4	2022-12-25 21:59:06	negative	active	angry
69	4	2022-12-25 21:59:11	negative	active	angry
70	4	2022-12-25 21:59:17	negative	active	angry
71	4	2022-12-25 21:59:22	negative	passive	sad
72	4	2022-12-25 21:59:28	negative	active	sad
73	4	2022-12-25 21:59:33	positive	active	happy
74	4	2022-12-25 21:59:39	positive	active	happy
75	4	2022-12-25 21:59:44	positive	active	happy
76	4	2022-12-25 21:59:50	positive	active	happy
77	4	2022-12-25 21:59:55	positive	active	happy

Joonis 22. Tabel *teacher_emotion*.

4 Tulemuste analüüs ja järeldused

Selle töö eesmärgiks oli luua näoilmete põhjal emotsiooni tuvastamismudeleid, mida saab edasi kasutada rakenduste loomiseks. Kasutati erinevaid klassifitseerimisalgoritme, nii klassikalisi kui ka närvivõrke. Närvivõrgud näitasid parimaid tulemusi mitmeklassilise ja binaarse klassifikatsiooni jaoks. Emotsioonid ei jagunenud töös mitte ainult 7 tüübiks, vaid ka passiivseteks/aktiivseteks ja positiivseteks/negatiivseteks.

Treening- ja testimisandmete jaoks kasutati andmestikku FER2013. Tuvastamistäpsuse parandamiseks kasutati mudelite loomisel Pythoni teeki, nagu *DeepFace* ja *face_recognition*, mille abil valiti pildid nii, et 7 emotsiooni määramise treeningmudelite andmestik oleks tasakaalukam.

Samuti muudeti piltide suurst 48x48-lt 24x24-le. Seda tehti, et mudelite loomise algoritmid töötaksid kiiremini. Klassikalised algoritmid hüperparameetrite valimisel mudelite loomiseks 48x48 suurusega piltide kasutamisel töötasid kuni 8 tundi, kasutades paralleelseid protsesse. Närvivõrke mudelite loomiseks kasutati Colab'i võimalusi: GPU — graafikaprotsessor, mis töötleb andmeid kiiremini, kuna ülesandeid täidetakse paralleelselt, mitte järjestikku nagu CPU korral.

Kuna närvivõrgud näitasid parimat tulemust, muudeti andmestikku: piltide suurused taastati 48x48-le. Järgmise sammuna närvivõrkude mudelite kvaliteedi parandamiseks on plaanis luua tasakaalustatud andmestik uute piltide genereerimise teel, kasutades generatiivseid võistlusvõrke (GAN - *Generative Adversarial Networks*).

Selles etapis loodi veebirakenduse prototüüp, mis ei ole veel käivitatud realses tunnis, et kontrollida rakenduse täielikku funktsionaalsust. Kuid prototüübi testimine näitas, et on võimalik reaalajas emotsioone tuvastada ja andmebaasi salvestada.

Õpetamisel ja õppimisel emotsionaalse seisundi uurimiseks on kõige olulisemad aktiivne/passiivne (*active/passive*) ja positiivse/negatiivse (*positive/negative*) tüüpi emotsioonid. Nende klassifikatsioonide jaoks loodud mudelid on näidanud head emotsioonituvastuse täpsust.

Täpsema emotsioonianalüüsi saamiseks saab luua ja integreerida oma rakendusse kõneemotsiooni tuvastamise mudelid *Speech emotion recognition (SER)*.

Samuti tuleb reaalseks veebirakenduseks luua tagasiside õppejõududele, tuginedes emotsioonidele. See peaks olema tundides saadud andmete visualiseerimine ja statistiline andmeanalüüs.

5 Kokkuvõte

Käesoleva magistritöö eesmärgiks oli luua masinõppel põhinev meetod üliõpilaste ja õppejõudude emotsionaalse seisundi jälgimiseks õppeprotsessi käigus näo emotsioonide järgi.

Mudelite treenimiseks ja testimiseks kasutati andmestikku Facial Expression Recognition 2013 (FER2013), mis sisaldab 35 887 pilti seitsme emotsioonitüübi (viha, rõõm, kurbus, üllatus, vastikus ja hirm pluss üks neutraalne) kohta.

Selle andmestikuga töötamisel ilmned järgmised probleemid: suur erinevus piltide arvus erinevates kategooriates (näiteks õnnelike emotsioonidega seotud kujutisi on 7215 treening- ja 1774 testimisandmetes, kuid vastikuse emotsioone on ainult 436 treening- ja 111 testimisandmetes vastavalt) ja mõne pildi kvaliteet on madal.

Teekide DeepFace (emotsioonide kokkulangevuse protsent) ja Face_recognition (näo tuvastamine pildil) abil valiti emotsioonide tuvastamise mõttes parimad pildid treening- (8140) ja testimisandmete jaoks (4577).

Käesolevas magistritöös on käsitletud 3 tüüpi emotsioonide klassifikatsiooni: klassifitseerimine 7 emotsiooni klassiks: (üllatus, viha, hirm, kurbus, neutraalne, rõõm, vastikus); klassifitseerimine 2 emotsiooni klassiks: positiivsed (rõõm) ja negatiivsed (viha, hirm, kurbus, vastikus) emotsioonid; klassifitseerimine 2 emotsiooni klassiks: aktiivsed (viha, vastik, hirm, rõõm, üllatus) ja passiivsed (kurbus) emotsioonid.

Mudelite koostamiseks kasutatakse klassikalisi ja sügavõppe klassifitseerimisalgoritme nagu SVM, Random Forest ja närvivõrke (CNN ja VVG-16). Realiseerimiseks kasutatakse Pythonit ning mudelite koostamiseks kasutati teke Scikit-learn ja Keras.

Emotsioonide määratlemiseks õpetamisel ja õppimisel on kõige olulisemad autori arvates passiivsed/aktiivsed ja positiivsed/negatiivsed, sest need võivad näidata selgemalt üliõpilaste huvitatust õppeprotsessist.

Eksperimentide tulemusena olid parimad näitajad närvivõrkudel:

- klassifitseerimine 7 emotsioonitüübiks – täpsus 0.72 (CNN)
- passiivsed/aktiivsed – täpsus 0.88 (VGG16)

- positiivsed/negatiivsed – täpsus 0.96 (VGG16).

Töö tulemuseks ei olnud mitte ainult mudelite loomine, vaid ka veebirakenduse prototüübi loomine Streamlit teegi abil, mis salvestab kõigi 3 klassifikatsioonitüübi emotsioonid reaajas PostgreSQL andmebaasi.

Peatükis 4. Tulemuste analüüs ja järeldused esitab autor lahendusi, kuidas parandada mudelite täpsust, näiteks kasutada generatiivseid võistlusvõrke (*GAN - Generative Adversarial Networks*) ja lisada kõneemotsiooni tuvastamise mudelid *Speech motion detection (SER)*. Samuti autor planeerib arendada veebirakenduse funktsionaalsust: lisada visualiseerimist ja statistilist andmeanalüüsi.

Kasutatud kirjandus

- [1] Ekman, P.: Basic emotions. In: T. Dalgleish, M. J. Power (eds.) Handbook of cognition and emotion, pp. 45–60. John Wiley & Sons Ltd. (1999).
<https://doi.org/10.1002/0470013494.ch3>
- [2] Ekman, P.: A methodological discussion of nonverbal behavior. *The Journal of Psychology*. 43(1), 141–149 (1957). doi: 10.1080/00223980.1957.9713059
- [3] Russell, J.: A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178 (1980). <https://doi.org/10.1037/h0077714>
- [4] Scherer, K.: What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729 (2005). doi:10.1177/0539018405058216
- [5] Haq, S., Jackson, P.: Multimodal Emotion Recognition. In W. Wang (ed.) *Machine Audition: Principles, Algorithms and Systems*, pp. 398-423. IGI Global (2011).
<https://doi.org/10.4018/978-1-61520-919-4.ch017>
- [6] Liu, S., Wang, W.: The Application Study of Learner's Face Detection and Location in the Teaching Network System Based on Emotion Recognition. In: *IEEE 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing*, vol. 1, pp. 394-397. IEEE Computer Society (2010). doi: 10.1109/NSWCTC.2010.97.
- [7] Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., Movellan, J.: The faces of engagement: Automatic recognition of student engagement from facial expressions. In: *IEEE Transactions on Affective Computing*, 5(1), 86-98 (2014). doi: 10.1109/TAFFC.2014.2316163
- [8] Ayvaz, U., Gürüler, H., Devrim, M.: Use of Facial Emotion Recognition in E-Learning Systems. *Information Technologies and Learning Tools*, 60(4), 95-104 (2017).
<https://doi.org/10.33407/itlt.v60i4.1743>
- [9] Lasri, I., Solh, A., Belkacemi, M.: Facial Emotion Recognition of Students using Convolutional Neural Network. *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)*, pp. 1-6 (2019). doi: 10.1109/ICDS47004.2019.8942386
- [10] Sathik, M., Jonathan, S.: Effect of facial expressions on student's comprehension recognition in virtual educational environments. *SpringerPlus* 2, 455 (2013).
<https://doi.org/10.1186/2193-1801-2-455>
- [11] Natalja Ivleva; Avar Pentel; Olga Dunajeva; Valeria Juštšenko (2022). *Machine Learning Based Emotion Recognition in a Digital Learning Environment*. Lecture Notes in Networks and Systems: International Conference on Interactive Collaborative Learning. "Learning in the Age of Digital and Green Transition". Vienna, Austria 27-30 September 2022. Lecture Notes in Networks and Systems: Springer Nature [ilmumas].
- [12] FER-2013, Available: <https://www.kaggle.com/datasets/msambare/fer2013>, [Kasutatud 12 2022]

- [13] Serengil, S., Ozpinar, A.: LightFace: A Hybrid Deep Face Recognition Framework. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 23-27. IEEE (2020). doi: 10.1109/ASYU50717.2020.9259802
- [14] Enrique Munoz-de-Escalona, José Juan Canas. Measuring of Available Resources. H-Workload 2017: The first international symposium on human mental workload, Dublin Institute of Technology. <https://doi.org/10.21427/D7DK96>
- [15] Rohan Gupta. Breaking Down Facial Recognition: The Viola-Jones Algorithm 07.08.2019. [Võrgumaterjal] Available: <https://towardsdatascience.com/the-intuition-behind-facial-detection-the-viola-jones-algorithm-29d9106b6999> [Kasutatud 12 2022]
- [16] Multiclass Classification Using Support Vector Machines. 11.11.2022. [Võrgumaterjal] Available: <https://www.baeldung.com/cs/svm-multiclass-classification> [Kasutatud 12 2022]
- [17] Convolutional Neural Network. Learn Convolutional Neural Network from... | by dshahid380 | Towards Data Science 24.02.2019. [Võrgumaterjal] [Kasutatud 12 2022]
- [18] Rohini G. Everything you need to know about VGG16. 23.09.2021. [Võrgumaterjal] Available: <https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918> [Kasutatud 12 2022]
- [19] Губко Павел, Горчаков Алексей, Буркина Мария. Метрики классификации и регрессии. [Võrgumaterjal] Available: <https://academy.yandex.ru/handbook/ml/article/metriki-klassifikacii-i-regressii> [Kasutatud 12 2022]
- [20] Project description DeepFace. [Võrgumaterjal] Available: <https://pypi.org/project/deepface/> [Kasutatud 12 2022]
- [21] Carlos Domínguez García. Visualizing the effect of hyperparameters on Support Vector Machines. 07.02.2021. [Võrgumaterjal] Available: <https://towardsdatascience.com/visualizing-the-effect-of-hyperparameters-on-support-vector-machines-b9eef6f7357b> [Kasutatud 12 2022]
- [22] C and Gamma in SVM. 17.12.2018. [Võrgumaterjal] Available: <https://medium.com/@myselfaman12345/c-and-gamma-in-svm-e6cee48626be> [Kasutatud 12 2022]
- [23] Keras documentation. EarlyStopping. [Võrgumaterjal] Available: https://keras.io/api/callbacks/early_stopping/ [Kasutatud 12 2022]

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Natalja Ivleva

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose "Masinõppel põhinev näo 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.

04.01.2023

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 loomingu tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtjaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.