



Sademetete lühiajaline prognoos masinõppe abil

Bakalaureusetöö

Üliõpilane: Marili Arus

Üliõpilaskood: 206542YAFB

Juhendaja: Sander Rikka, vanemteadur, TalTech meresüsteemide instituut

Kaasjuhendaja: Aarne Männik, vanemteadur, TalTech meresüsteemide instituut

Õppekava: Rakendusfüüsika

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: Marili Arus

Töö vastab bakalaureusetööle esitatavatele nõuetele.

Juhendaja: Sander Rikka,

Kaasjuhendaja: Aarne Männik

Annotatsioon

Käesolev töö käsitleb lühiajalist sademete prognoosimist süvanärvivõrgu meetodiga. Töö eesmärgiks oli luua punktmõõtmiste andmetega sademete kvalitatiivse prognoosi mudeli prototüüp kasutades LSTM süvanärvivõrku.

Töös kasutati Tallinn-Harku aeroloogiajaama tunnipõhiseid andmeid aasta 2011 algusest aasta 2023 lõpuni. Andmetes olid esindatud aeg, sademed, tuul, õhutemperatuur, õhuniiskus, õhurõhk ja pilved. Andmed puhastati, standardiseeriti, ning koostati libiseva aknaga 15 tunnised valimid, kusjuures 12 tundi on kõiki parameetreid sisaldavad mineviku andmed, ning 3 tundi on prognoositavad sademe klassi (sajab/ei saja) väärtused. LSTM mudeli hüperparameetrite optimeerimiseks rakendati Optuna teeki. Optimeeritud mudeli hindamiseks kasutati tuvastustõenäosust (POD), valehäirete osakaalu (FAR) ning õiglast tabamismäära (ETS). Mudeli meetrikaid võrreldi püsivusprognoosiga ning meetrikate halvenemist võrreldi teiste masinõppe mudelitega (ConvLSTM, PredRNN ning PredRNN++).

Tulemused näitasid, et loodud prototüüp lisas püsivusprognoosile tagasihoidlikku, kuid pidevat väärtust igal ajasammul. Samuti tuli välja, et LSTM tüüpi mudelite meetrikate väärtused halvenesid reeglina aeglasemalt, kui keerulisemad PredRNN tüüpi mudelite omad, erandiks oli käesoleva töö LSTM mudeli FAR, mis halvenes kiiremini, kui PredRNN++ mudelil. Parima mudeli meetrikate väärtused 1. tunnil olid $POD = 0,70$, $FAR = 0,30$ ning $ETS = 0,49$.

Arutelus pakutakse lisaks ideid LSTM prognoosimudeli võimalikuks parandamiseks.

Abstract

This thesis addresses precipitation nowcasting using deep-neural network methods. The objective was to develop a prototype of a qualitative precipitation nowcasting model using an LSTM deep neural network relying mainly on in situ measurements.

Hourly data from the Tallinn-Harku aerological station covering the period from the beginning of 2011 to the end of 2023 was used in the study. The dataset included time, precipitation, wind, air temperature, relative humidity, air pressure, and clouds. The data was cleaned, standardized and made into 15 hour samples using a sliding window; the first 12 hours consisted of past data for all parameters and the subsequent 3 hours represented the forecast precipitation class values (rain/no rain). To optimize the LSTM model, the Opuna library was used. Model performance was evaluated using the probability of detection (POD), false alarm ratio (FAR), and equitable threat score (ETS). The model metrics were compared with the persistence forecast, and the deterioration of the metrics was further compared with other types of machine learning models (ConvLSTM, PredRNN, and PredRNN++) from previous studies.

The results showed that the developed prototype added modest but consistent value over the persistence forecast at each time step. It was also found that the performance metrics of LSTM-type models generally degraded more slowly than those of the more complex PredRNN-type models. An expectation was the FAR of the LSTM model developed in this study, which deteriorated faster than that of the PredRNN++ model. The metric values of the best model were POD = 0,70, FAR = 0,30 and ETS = 0,49.

In the discussion, additional ideas are proposed for potential improvements to the LSTM forecasting model.

Lühendite ja mõistete sõnastik

Andmete assimilatsioon	<i>numbrilise modelleerimise haru, mis kasutab tehnikaid vaatluste ja mudelite vaheliste vigade (erinevuste) vähendamiseks [1].</i>
ConvoLSTM	<i>Convolutional Long Short-Term Memory</i> Konvolutsiooniline pikk lühimälu
ETS	<i>Equitable Threat Score</i> Õiglane tabamismäär
FAR	<i>False Alarm Ratio</i> Valehäirete osakaal
LSTM	<i>Long Short-Term Memory</i> Pikk lühimälu
POD	<i>Probability of Detection</i> Tuvastustõenäosus
PPI	<i>Plan Position Indicator</i> Kiire tasapinna nurga indikaator <i>mõõtmistulemused, mis on saadud maapinnale projitseeritud ühe fikseeritud radarikiire tõusunurgal [2]</i>
PredRNN	<i>Predictive Recurrent Neural Network</i> Ennustav korduv närvivõrk
TS	<i>Threat score</i> Tabamismäär

Sisukord

1	Sissejuhatus	6
2	Teoreetiline taust.....	7
2.1	Sademetes ülevaade.....	8
3	Metoodika	10
3.1	Andmete valimine ja töötlemine	10
3.1.1	Andmete valimine.....	10
3.1.2	Andmete töötlemine	10
3.1.3	Valimite moodustamine	11
3.2	Sademetes seosed teiste suurustega	12
3.3	LSTM mudeli arhitektuur ja optimeerimine	13
3.4	LSTM mudeli hindamine.....	14
3.4.1	Püsivusprognosis	14
3.4.2	Hindamismeetrika	15
4	Tulemused.....	17
4.1	LSTM mudeli ning püsivusprognosis ennustuste tulemused ja meetrikad .	17
4.2	LSTM mudeli ja püsivusprognosis võrdlus	18
4.3	Võrdlus uuringutega.....	19
4.3.1	PredRNN++ kasutades radaripilte	19
4.3.2	PredRNN ning ConvLSTM kasutades kohalike sademetes andmeid ...	20
5	Arutelu	22
	Kasutatud kirjandus	23
	Lisad	26

1 Sissejuhatus

Ootamatu vihmased on igapäevane probleem. See häirib spordi tegemist, välisõhu üritusi, ehitust ja muud. Mõnda tüüpi sademed kujunevad ja muutuvad kiiresti (minutid kuni tunnid), mõjutades seejuures väikest piirkonda. Traditsiooniliste numbriliste ilmaprediktsioonide ennustamiseks kasutatavate andmete kogus on tohutu, mistõttu tehakse uusi ennustusi vaid mõni kord päevas. Selle aja jooksul võib juba olnud tekkinud sajupilv, ning ka sadu alanud. Lisaks võib selliste prediktsioonide lahutusvõime jääda lokaalsete sademete jaoks liiga väikseks. On vaja lihtsamaid ja kiiremaid meetodeid ennustamiseks: lühiajaline prognoos ehk *nowcasting*.

Masinõppe, sh süvaõppe, populaarsuse kasv on toimunud ka ilmaennustuse valdkonnas. Masinõppe prediktsioonide eelis on see, et isegi kui prediktsiooni treening võtab kaua aega, siis treenitud prediktsiooni ennustus on kordades kiirem kui traditsiooniline numbriline prediktsioon. Suurem osa süvaõppe ja *nowcasting* prediktsioonide kasutamiseks satelliiti või radarit; prediktsioonid mis õpivad vaid maapealsetelt mõõtmistelt on kirjanduses vähem esindatud.

Töö eesmärgiks on süvanärvivõrgu meetoditega luua sademete lühiajalise kvalitatiivse prognoosi LSTM (*Long Short-Term Memory* ehk pikk lühimälu) prediktsiooni prototüüp, kasutades vaid kohalikke mõõtmisi, see tähendab ilma kaugseire andmeteta. Selline prediktsioon peab toetuma vaid lokaalselt mõõdetud füüsikaliste suuruste vahelisele seostele ning nende muutustele ajas.

Kuna töös kasutatavad parameetrid vajavad mõõtmiseks pigem väikseid sensoreid (suurim on pilvede mõõtmiseks vajalik seilomeeter), siis selline lähenemine võimaldaks luua kaasaskantava sensorite komplekti, mille saaks vajaduspõhiselt ümber liigutada. Näiteks oleks võimalik panna selline kompleks püsti Tallinna lauluväljakule ürituste eelselt sadu ennustama, või saaks ehitusfirma viia selle komplekti erinevatele objektidele.

2 Teoreetiline taust

Ilmaennustuses laialdaselt kasutusel olevad numbrilised mudelid ei anna lühiajalisel prognoosimisel parimaid tulemusi. Selle peamine põhjus on prognooside tegemisele kuluv aeg: kõigepealt kulub mitmeid minuteid vajalike andmete kohale jõudmiseks [3], ning seejärel mõni tund arvutuste stabiliseerumiseks [4][5]. Selleks ajaks on ennustatav periood juba möödunud ning võib-olla sadama hakanud – hoovihmad võivad konvektsiooni teel tekkida minutitega [6]. Teine probleem on nende väike lahutusvõime. Sellised mudelid on paremini kohandatud ulatuslike sademete ennustamiseks, ning lokaalsed hoovihmad, mida esineb palju suvel, jäävad neil märkamata [7][5]. Lahutusvõime suurendamine pole efektiivne lahendus, sest selle tegemisega pikeneb prognoosi tegemise aeg eksponentsiaalselt [7].

Lühiajaline prognoos ehk *nowcasting* on ilmaennustuse meetod, mille eesmärk on prognoosida ilma 6 tunni ulatuses [7]. See on eriti oluline nii ohtlike loodusnähtuste – tornaadod, äkilised üleujutused, kiiresti kujunevad äikesetormid – kui ka väiksemate ohtude, nagu jäävihm või tihe udu, varajaseks märkamiseks [7]. Lühiajalise prognoosi tehnikad toetuvad tavaliselt maapealsetele mõõtmistele, radarile ning satelliidi andmetele [7], ning need mudelid on traditsioonilistest numbrilistest mudelitest lihtsamad, mis laseb neil lühikesel ajaskaalal efektiivselt töötada [3]. Küll aga langeb lühiajaliste prognooside täpsus palju kiiremini kui traditsioonilistel numbrilistel mudelitel, ning neid on mõistlik kasutada 6 tunni piires [8].

Peamised sademete lühiprognosi tehnikad põhinevad ekstrapoleerimisel. See käsitleb sademeid kui sademete välja, mis ajas edasi liigub ning deformeerub [3]. Meetod tugineb Lagrangiane'i püsivuse eeldusele: sademete välja sisemine struktuur on konstantne ning kõik muutused on tingitud suurte õhuvooludega kaasaliikumise tõttu [8]. See eeldus on lihtsustus, mis töötab kõige paremini lokaalselt, kuid siiski erineb sageli palju tegelikkusest [8]. Üks probleemidest seisneb sademete välja kasvamises ning hääbumises [9][3]. On proovitud erinevaid lähenemisi selle meetodi täpsemaks tegemiseks, kuid pole leitud üldist lahendust, mis ei kaasaks kulukat andmete assimilatsiooni¹ [8]. Sellest tulenevalt on hakanud uurima masinõpet kui võimalikku lahendust sellele probleemile [8].

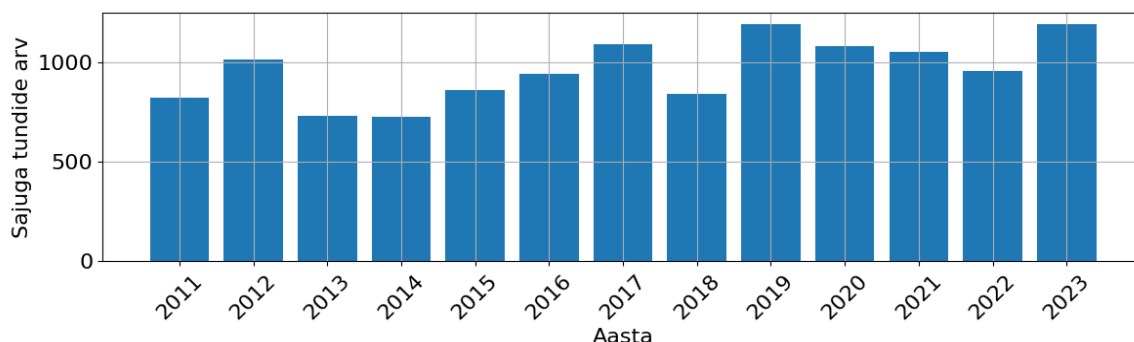
On tõestatud, et süvaõppe mudelid on radari piltide ekstrapoleerimisel oskuslikumad, kui traditsioonilised Lagrangiane'i eeldusel põhinevad lühiprognosi mudelid [10]. Süvaõppe mudelid saavad efektiivselt kasutada sageli värskendatavaid andmeid radaritelt, satelliitidelt ning teistelt vaatlussüsteemidelt [10]. Nad saavad hästi hakkama mitte-lineaarsete nähtuste, nagu hoovihmad, kirjeldamisega – miski mida on numbriliste meetodite ning ekstrapolatsiooniga raske teha [10]. Eeliseks on ka kiirus:

¹ Andmete assimilatsioon on numbrilise modelleerimise haru, mis kasutab tehnikaid vaatluste ja mudelite vaheliste vigade (erinevuste) vähendamiseks [1].

treenitud süvaõppe mudelil kulub superarvuteid kasutamata prognoosimiseks sekundeid [10].

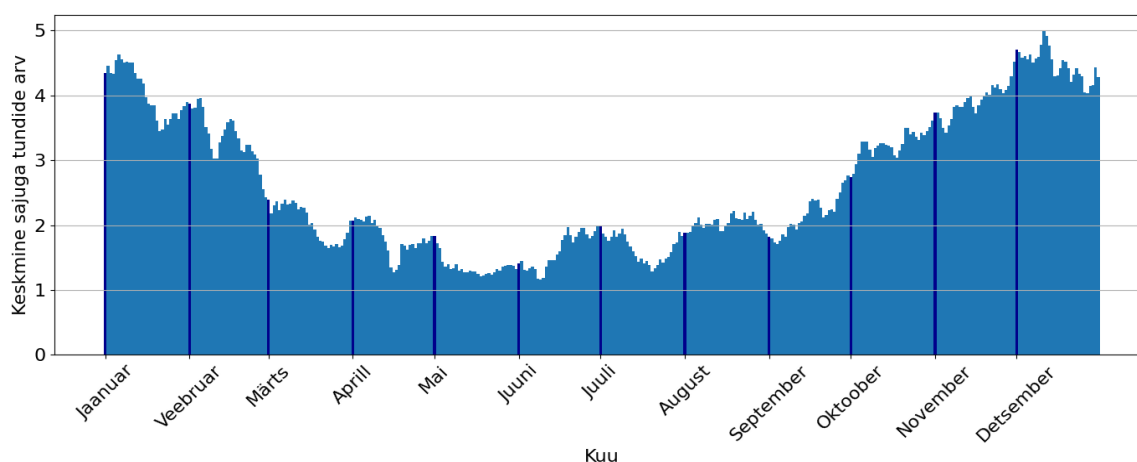
2.1 Sademete ülevaade

Iga aasta summaarne sajuga tundide arv on toodud joonisel 1. On näha, et sademete sagedused võivad aastast aastasse palju erineda. Kõige vähem sajutunde (727) oli aastal 2014, mil sajutunnid moodustasid 8,3% kõigist tundidest, ning kõige rohkem (1192) aastal 2019, mil sajutunnid moodustasid 13,6% kõigist tundidest; keskmiselt sajab 962.5 tundi aastas. Terve perioodi jooksul sadas keskmiselt 11,2%-l tundidest.



Joonis 1: Tallinn-Harku aeroloogiajaama sajuga tundide arv aasta kaupa 2011-2023

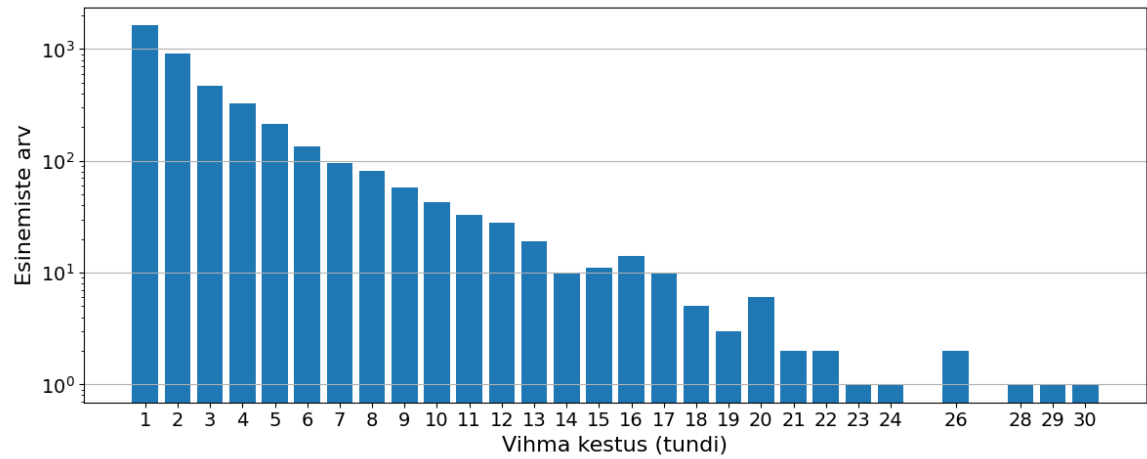
Joonisel 2 on histogramm silutud keskmisest sajuga tundide arvust päeva kaupa; kuude esimesed päevad on tehtud tumesiniseks. Andmete silumiseks kasutati liikuva keskmise meetodit aknapikkusega 15 päeva. On näha sujuvat üleminekut hooajast hooaega. Kõige väiksem sajutundide osakaal oli mais 5,7%-ga ning kõige suurem detsembris, millal sadas 18,9% tundidest.



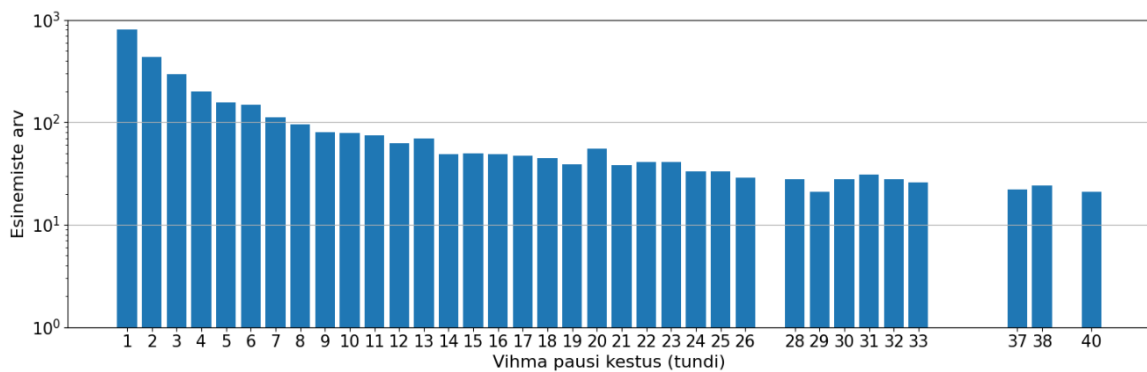
Joonis 2: Tallinn-Harku aeroloogiajaama 2011-2023. aasta keskmine sajuga tundide arv päeva kaupa, silutud

Joonisel 3 on toodud perioodi 2011-2023 vihma kestvuste histogramm. Enamik sademeperioode olid lühikesed: 39,5% sademe perioodidest kestis maksimaalselt ühe tunni, 22,3% kestis maksimaalselt 2 tundi. Viis korda kestis sadu rohkem kui

ööpäev. Saju pausidel (sajuga tundide vaheline aeg) on samamoodi geomeetriline jaotus (joonis 4), kuid see on palju raskema sabaga: 19,4% saju pausidest kestis vähemalt ühe, kuid vähem kui kaks tundi, ning kõige pikem vihma paus oli 535 tundi ehk üle 22 päeva



Joonis 3: Perioodi 2011-2023 vihma kestvuste histogramm



Joonis 4: Perioodi 2011-2023 vihmapiuside kestvuste histogramm ($n > 20$)

3 Metoodika

3.1 Andmete valimine ja töötlemine

3.1.1 Andmete valimine

Käesoleva töö andmed saadi Eesti Keskkonnaagentuuril [11] avaandmete portaalist Tallinn-Harku aeroloogiajaamast. Valiti 1 tunni mõõtmise intervalliga andmed. Andmete periood oli 2011. a algusest kuni 2023. a lõpuni.

Valiti järgmised füüsikalised suurused:

- Ühe tunni sademete summa, mm
- Ühe tunni maksimaalne tuule kiirus, m/s
- Viimase 10 minuti keskmine tuule kiirus, m/s
- Viimase 10 minuti valdav tuule suund, °
- Ühe tunni maksimaalne õhutemperatuur, °C
- Ühe tunni minimaalne õhutemperatuur, °C
- Õhutemperatuur täistunnil (hetkväärtus), °C
- Suhteline õhuniiskus täistunnil (hetkväärtus), %
- Õhurõhk merepinnal täistunnil (hetkväärtus), hPa
- Alumiste pilvede kõrgus, m
- Alumiste pilvede hulk, (0, 8)

Täiendavalt lisati ka:

- Tund, (0, 23) h
- Päev aastas, (0, 364) või (0, 365) d

Aasta ja päev on tsüklilised: 1. jaanuar ja 31. detsember, nagu ka kell 00:00 ja kell 23:00, on üksteisega kõrvuti. Arvuliselt aga on 0 ja 364 või 0 ja 23 lõigu otstes, ning üksteisest väga erinevad. Selleks, et anda LSTM mudelile edasi nende tsükliline iseloom, võeti ajalistest suurustest lisaks $\sin(2\pi * \frac{x}{n})$ ja $\cos(2\pi * \frac{x}{n})$, kus x tähistab muutuvat tunni/päeva väärtust, ning n maksimaalset tunni/päeva väärtust. Siinus-koosinus paar kannab päeva ja tunni väärtused ühikringile.

3.1.2 Andmete töötlemine

Seilomeeteri mõõdetud pilvede andmed olid jagatud nelja kihti. Iga kihi kõrguse piirang oli seilomeetri maksimaalne mõõdetav kõrgus: esimeses kihis oli sellel tunnil kõige madalam mõõdetud kiht, mis võis olla ka 7,6 km kõrgusel kui sellest allpool pilvi ei olnud [12].

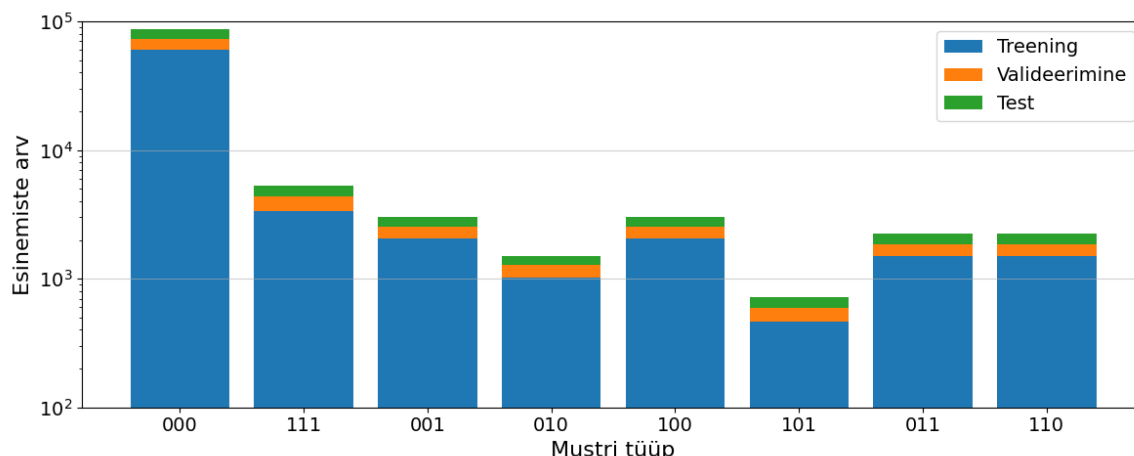
Käesolevas töös kasutati vaid kõige alumise kihi andmeid, sest sajupilvede baasid on reeglina alla 2 km kõrgusel [6]. Pilvede kõrguse andmetes võeti maksimum kõrguseks 2000 m: kõrguse väärtused suuremad kui 2000 m võeti võrdseks 2000 m-iga. Grupeerides kõrged mittesajupilved, keskendus LSTM mudel rohkem sajupilvedele. Tühja taeva (pilvede kogus = 0) puhul pandi puuduva kõrguse asemele -1. Andmed järjestati aja järgi ning pandi andmeraamistikku; sademete mm väärtus tehti kahendväärtuseks (0 – ei saja, 1 – sajab). Kuna andmestik oli suur ning puuduvate andmetega ridu oli vaid 1,65%, siis need read eemaldati. Alles jäi 112 071 tunni andmed.

3.1.3 Valimite moodustamine

Enne valimite moodustamist andmed standardiseeriti, et suurte väärtustega füüsikalised suurused ei domineeriks. Selleks kasutati klassi *sklearn.preprocessing.StandardScaler* [13], mis viib iga füüsikalise suuruse jaotuse kujule, millel on keskvärtus 0 ja standardhälve 1. Standardiseerimisest jäeti välja sademed, pilvede hulk, ning siinus ja koosinus päevast ning tunnist, sest standardiseerimine oleks rikkunud nende kategoorilist või tsüklilist infot, ning ükski neist ei olnud domineerivate väärtustega.

Käesolevas töös ennustati 12 mineviku tunni põhjal 3 tuleviku tundi; selleks libises üle originaal andmestiku 15 tunnine aken. Kõigepealt kontrolliti kas akna sees on ajalisi hüppeid. Kui hüpe esines, libises aken ühe tunni võrra edasi ning kontrolliti uuesti. Hüpete puudumisel lisati 12 esimest tundi sisendite massiivi, ning 3 viimast tundi väljundite massiivi, kusjuures sisendite massiivis olid olemas kõik füüsikalised suurused ja väljundite massiivis vaid sademete väärtused. Iga 12 + 3 h paar moodustas ühe valimi. Kokku oli 104 186 valimit.

Joonisel 5 on toodud valimite tuleviku tundide sajab/ei saja (1/0) mustrite histogramm. Täielikult sajuta muster '000' moodustab 82,4% kõigist mustritest. Sadu sisaldavatest mustritest on '111' kõige sagedasem: kõigist mustritest moodustab see 5,3%, ning sajuga mustritest 30,1%. Teisel kohal sajuga mustritest on '001', mida on peaaegu poole vähem, kui mustrit '111'. Kõige vähem esineb mustrit '101', mis moodustab 0,7% kõigist mustritest, ning 4,0% sajuga mustritest.

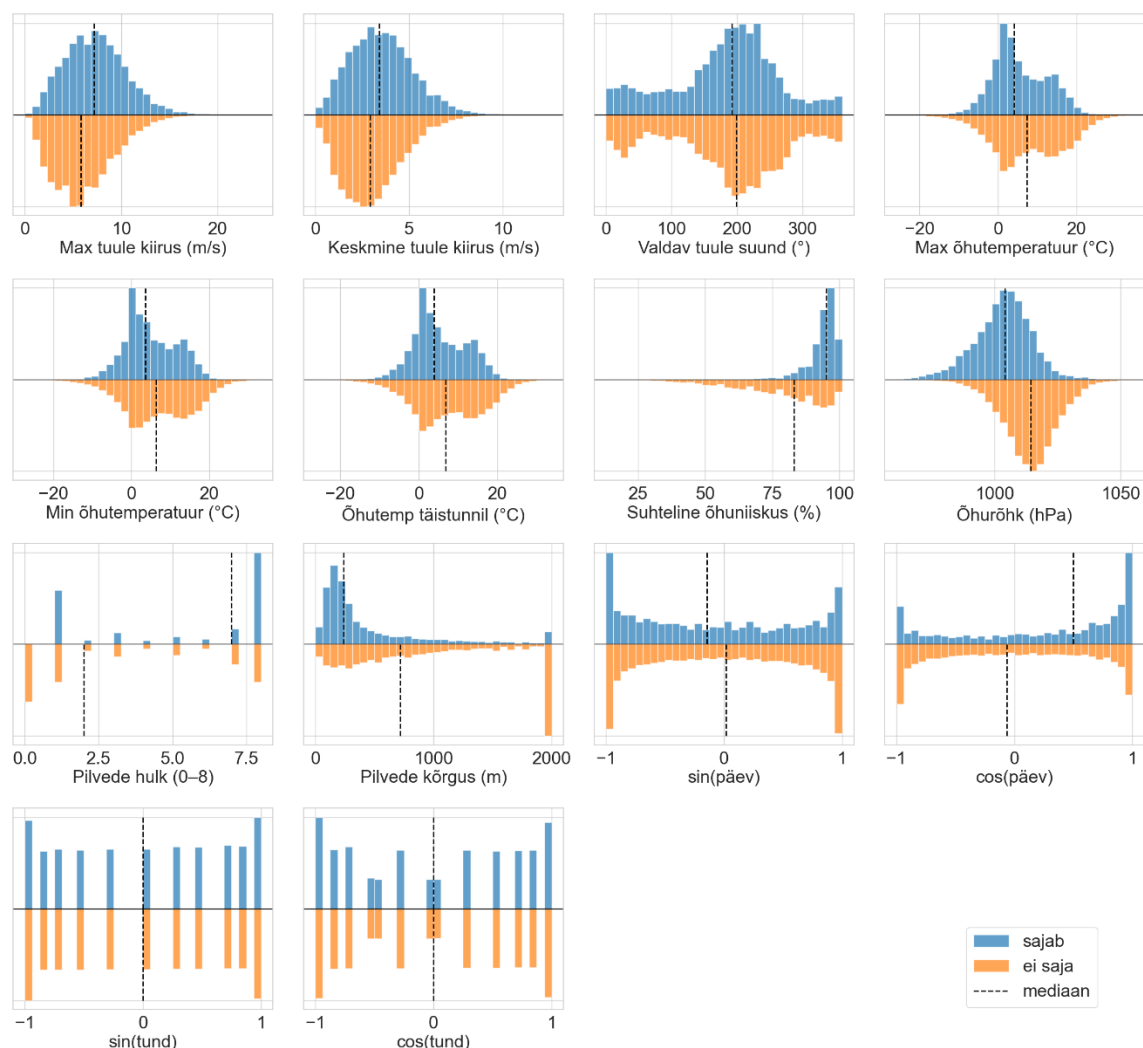


Joonis 5: Valimite tuleviku tundide (3h) mustrite histogramm

Andmed jagati treening-, valideerimis- ning testandmeteks ajalist järjekorda rikkumata. Jaotus oli järgmine: LSTM mudel treenis umbkaudu 2011-2019 andmetel, valideeris 2020-2021 andmetel, ning testis 2022-2023 andmetel. Ajalise järjekorra säilitamine oli oluline, sest valimid olid tehtud libiseva aknaga, ning andmeid segades satuks osaliselt identseid valimeid nii treening kui ka testandmetesse. Mustrite jaotumist treening-, valideerimis- ning testandmete vahel on näha joonisel 5.

3.2 Sademete seosed teiste suurustega

Joonisel 6 on füüsikaliste suuruste normaliseeritud histogrammid vastavalt sajab / ei saja klassile; musta punktiirjoonega on märgitud mediaanide väärtused. Joonise jaoks eemaldati pilvede kõrguste väärtustest -1, mida kasutati pilvede puudumise tähistamiseks. Suurimad erinevused klasside jaotuste vahel on suhtelisel õhuniiskusel, õhurõhul ja pilvede kõrgusel: sadu esineb kõrgema õhuniiskuse, madalama õhurõhuga ning madalatest pilvedest. On näha, et sademed esinevad natuke kõrgema tuulekiiruse, ning madalama õhutemperatuuriga. Loogiline seos on ka sademete ja pilvede hulga vahel: sademeid esineb suurema pilvede hulga korral. Aasta päeva trigonomeetrilistel väärtustel on samuti näha erinevusi saju ja mitte-saju klassi vahel. Tuule suuna histogrammid on peaaegu erinevusteta, ning tunni trigonomeetrilistel väärtustel erinevust ei esine. See ei tähenda veel, et need ei ole LSTM mudelile kasulikud. Joonisel on võrreldud iga suurust eraldi – need ei käsitle mitme parameetri koosmõju.



Joonis 6: Füüsikaliste suuruste normaliseeritud histogrammid vastavalt sajab / ei saja klassile

3.3 LSTM mudeli arhitektuur ja optimeerimine

Nagu ka andmete töötlemine, tehti LSTM mudeli treening ja optimeerimine programmeerimiskeeles Python integreeritud programmeerimiskeskonnas PyCharm. Peamiselt kasutati teke Pandas, NumPy, scikit-learn, TensorFlow, Keras, Seaborn ja Matplotlib [14][15][13][16][17][18][19]. Koodi kirjutamisel ja vigade leidmisel kasutati tehisintellekti tekstirobotite ChatGPT ja Copilot abi [20][21].

Mudel koosnes sisendkihist, mis võttis vastu sisendi kujuga 12 rida ja 15 tulpa, kahest järjestikkusest peidetud LSTM kihist ning väljundkihist mis väljastas ennustuse kujuga 3 rida ja 1 tulp.

Mudeli hüperparameetrite tuunimiseks kasutati Optuna teeki [22]. Kokku tuuniti kaheksat hüperparameetrit: mõlema LSTM kihi neuronite arv, väljalangemise määr (*dropout rate*), rekurrentne väljalangemise määr (*recurrent dropout rate*), õppimiskiirus, partii suurus, gamma ja alfa. Väljalangemise määra parameetreid

kasutatakse mudeli treeningandmetel üleõppimise ärahoidmiseks [23]. Gamma ja alfa on *BinaryFocalCrossentropy* kaofunktsiooni (*loss function*) parameetrid, mis aitavad saju ja mitte-saju klasse võrdsustada: gamma > 0 puhul keskendub mudel rohkem raskemini ennustatavatele valimitele ning alfa > 0.5 teeb saju klassi kaalu suuremaks [24]. Väljalangemise määrade, gamma ja alfa puhul jäeti Optunale võimalus need välja lülitada, st valida vastavalt väärtused 0, 0, 0 ja 0.5. Optimeerimise eesmärgiks seati maksimeerida õiglast tabamismäära (*Equitable Threat Score* ehk ETS).

Optimeerimiseks tehti 100 katset, kusjuures iga katsel treeniti mudelit maksimaalselt 40 epohhi. Treeningu kontrollimiseks ja kiirendamiseks pandi peale varajane peatamine: kui 7 epohhi pole validatsiooni kadu paranenud vähemalt 1e-5 võrra, treening lõppeb ning salvestatakse selle katse parimad hüperparameetrid.

Olles leidnud parimad hüperparameetrid, treeniti nendega lõplik LSTM mudel. Seekord pandi piiriks 100 epohhi, ning varajane peatumine sama minimaalse paranemisega 30 epohhi.

3.4 LSTM mudeli hindamine

3.4.1 Püsivusproгноos

LSTM mudeli kasulikkuse hindamiseks võrreldi seda püsivusproгноosiga (*persistence forecast*). See on üks lihtsamatest proгноosi meetoditest, mis ennustab alati, et järgmine päev või tund tuleb täpselt samasugune nagu praegune päev või tund. Püsivusproгноos on kõige usaldusväärsem lühikese aja jooksul, ning ennustuse täpsus väheneb märgatavalt, kui proгноosiperiood pikeneb. Konvektsioonist tingitud sademete püsivusproгноos hakkab halvenema juba 5ndast minutist ning väiksemad äikesetormid võivad juba 30 minuti jooksul palju muutuda [3].

Käesolevas töös koostati püsivusproгноos järgmiselt: iga valimi viimane, ehk 12nda tunni sademete väärtus võeti ennustuseks ehk järgneva kolme tunni väärtuseks.

LSTM Mudeli ja püsivusproгноosi meetrikate võrdluseks kasutati oskusskoori (*Skill Score*)

$$S = \frac{s_f - s_r}{s_p - s_r}$$

kus s_f on mudeli ennustuse meetrika väärtus, s_r on võrreldava ennustuse (püsiv-proгноosi) meetrika väärtus ning s_p on meetrika parim väärtus. Oskusskoori parim väärtus on 1, kusjuures 0 tähendab, et mudeli ennustus ei ole parem võrreldavast proгноosist [25].

Mudeli võrdlemiseks teiste uuringute mudelitega vaadati meetrikate halvenemist.

3.4.2 Hindamismeetrika

Alapeatükis 3.2.1 selgus, et sajuta tundide osakaal on oluliselt suurem kui sajuga tundide osakaal. Eeldades sajuga tundide suurimat osakaalu (detsembri keskmine 18,9%), saavutaks mudel täpsuse 0,81 ka siis, kui see ennustaks igal tunnil sajuta ilma. Seetõttu kasutati hindamismeetrikaid, mis keskenduvad vaid positiivsele klassile: tuvastustõenäosus (*Probability of Detection* ehk POD), valehäirete osakaal (*False Alarm Ratio* ehk FAR) ning õiglane tabamismäär (ETS).

LSTM Mudel tagastas iga tunni prognoosi tõenäosusena sündmuse esinemiseks. Tõeväärtuste (0 või 1) saamiseks tuli ennustusi kõigepealt teisendada. Selleks valiti läve väärtus, millest suurema tõenäosusega tunnid liigitati saju (1) ning väiksema tõenäosusega sajuta (0) klassi. Täpsete lävewäärtuste määramine on alapeatükis 4.1.

Kõiki meetrikate arvutamiseks tuli kõigepealt koostada sagedustabel ennustustest ja tegelikest väärtustest, kus 0 tähistas negatiivset klassi ehk saju puudumist, ning 1 positiivset ehk saju klassi. Tabelis 1 on toodud tüüpilise sagedustabeli struktuur; vale positiivne ja vale negatiivne on mudeli eksimused.

Tabel 1: Sagedustabel ennustustest ja tegelikest väärtustest

	Tegelik 1	Tegelik 0
Ennustatud 1	Õige Positiivne (ÕP)	Vale Positiivne (VP)
Ennustatud 0	Vale Negatiivne (VN)	Õige Negatiivne (ÕN)

POD näitab kui palju tegelikest sündmustest õigesti ennustati.

$$POD = \frac{\tilde{O}P}{\tilde{O}P + VN}$$

POD jääb lõiku [0, 1], parim tulemus on 1.

FAR näitab kui palju ennustatud sündmustest olid valed.

$$FAR = \frac{VP}{\tilde{O}P + VP}$$

FAR jääb lõiku [0, 1], parim tulemus on 0.

Õiglane tabamismäär hindab üleüldist sündmuse ennustamise oskust, karistades nii vale positiivseid kui ka vale negatiivseid. Õiglane tabamismäär on sarnane tavalisele

tabamismäärade (*Threat score* ehk TS), kuid on selgem hinnang mudeli oskusele, sest see eemaldab mõju, mis tuleks juhuslikust ennustamisest [26].

ETS arvutatakse järgmiselt:

$$ETS = \frac{\tilde{O}P - R}{\tilde{O}P + VP + VN - R}$$

kus

$$R = \frac{(\tilde{O}P + VP)(\tilde{O}P + VN)}{\tilde{O}P + VP + VN + \tilde{O}N}$$

on juhusliku ennustuse oodatud õigete positiivsete arv.

ETS jääb lõiku $[-\frac{1}{3}, 1]$, parim tulemus on 1, ning 0 näitab, et mudel polnud juhuslikust ennustusest parem, ehk mudeli prognoos on oskusetu [26].

Meetrika väärtuste halvenemise arvutamiseks kasutati valemit

$$H = \frac{a - l}{a} * 100\%$$

kus a on meetrika algväärtus ning l on meetrika lõppväärtus.

4 Tulemused

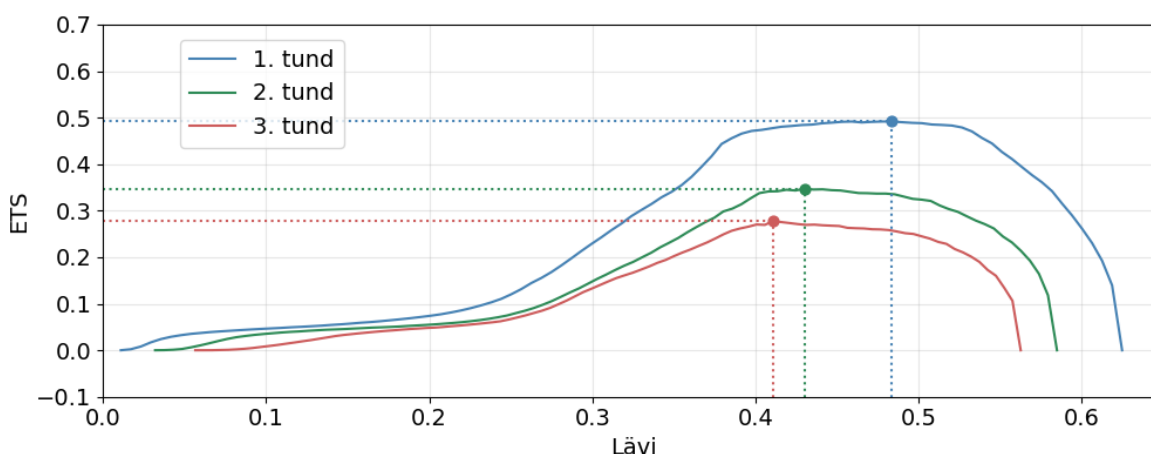
Optimeeritud LSTM mudeli täpne struktuur on leitav lisast 1. Parim tulemus saavutati 8ndal epohhil.

Optimeerimise käigus valiti parimaks positiivse klassi kaaluks 0,77, negatiivne klass sai kaaluks seega $1 - 0,77 = 0,23$. Algne klasside jaotus oli: sadu – 11,2% ning mitte-sadu – 88,7%, peale kaalude lisamist oli uus efektiivne klasside jaotus: sadu – 29,7% ning mitte-sadu – 70,3%.

4.1 LSTM mudeli ning püsivusproгноosi ennustuste tulemused ja meetrikad

LSTM mudeli tagastatud tõenäosuste väärtused ei ole kalibreeritud, kuid kuna eesmärgiks on kvalitatiivne, mitte tõenäosuslik prognoos, siis piisab vaid läve tuunimisest [27].

LSTM mudeli ETS arvutamiseks valiti kõigepealt iga tunni jaoks lävi, mis maksimeeris ETS väärtuse (joonis 7). Selleks leiti iga tunni prognoosi väärtuste vahemik ning arvutati ETS välja 100s võrdse vahega kohas; iga tunni maksimaalsele ETS väärtusele vastav koht valiti selle tunni läve väärtuseks. Joonisel on märgitud iga tunni jaoks ETS suurim väärtus. Parimad läve väärtused tundidele 1-3 olid vastavalt 0,48, 0,43 ning 0,41. POD ja FAR arvutati samuti nende lävede järgi (tabel 2). Tabelis 3 on toodud LSTM mudeli meetrikate halvenemised protsentides erinevate ajasammude kohta. Kõik meetrikad halvenevad esimesel ajasammul kõige rohkem; 2. tunnilt 3ndale minnes on kahjud väiksemad.



Joonis 7: LSTM mudeli ETS väärtus sõltuvalt läve valikust tundide kaupa

Tabelis 4 on toodud POD, FAR ja ETS väärtused püsivusproгноosi jaoks. Ühe tunni piires on meetrikad parima tulemusega võrreldes juba mõõdukalt langenud. Suhteliselt kõrge POD on tõenäoliselt tingitud sellest, et 12+3 tunni valimises on 3-tunniste sadu/mitte-sadu mustrite hulgas '111' kõige sagedamini esinev sadu

sisaldav muster (joonis 5). Kuna püsivprognoosi ennustused järgivad andmestiku jaotust, siis ennustatud VP-d ja VN-d on ligikaudu võrdsed, ning POD ja FAR summa on iga tunni jaoks 1.

Tabel 2: LSTM mudeli Läve, POD, FAR ja ETS väärtused tundide kaupa

	1. tund	2. tund	3. tund	Parim tulemus
POD	0,70	0,62	0,61	1
FAR	0,30	0,46	0,56	0
ETS	0,49	0,35	0,28	1
Lävi	0,48	0,43	0,41	

Tabel 3: LSTM mudeli meetrikate halvenemised (%) vastavalt ajasammudele

	1. → 2.	2. → 3.	1. → 3.
POD	11,4	1,6	12,9
FAR	53,3	21,7	86,7
ETS	28,6	20,0	42,9

Tabel 4: Püsivprognoosi POD, FAR ja ETS väärtused tundide kaupa

	1. tund	2. tund	3. tund	Parim tulemus
POD	0,69	0,56	0,48	1
FAR	0,31	0,44	0,52	0
ETS	0,48	0,33	0,26	1

4.2 LSTM mudeli ja püsivprognoosi võrdlus

Tabelis 5 on toodud mudeli meetrikate oskusskoorid võrreldes püsivprognoosiga. Mudeli oskus paistab kõige rohkem välja POD väärtuste võrdluses: igal ajasammul prognoosib LSTM mudel püsivprognoosist rohkem sademetega tunde õigesti, eriti domineerib see viimasel tunnil.

Tabel 5: Mudeli meetrikate oskusskoorid võrreldes püsivprognoosiga

	1. tund	2. tund	3. tund
S (POD)	0,03	0,14	0,26
S (FAR)	0,03	-0,04	-0,07
S (ETS)	0,02	0,02	0,03

Samas ilmneb FAR väärtuste põhjal, et parem POD väärtus saavutatakse osaliselt vale positiivsete arvu suurenemise arvelt. Teisel ja kolmandal tunnil on FAR oskusskoor negatiivne, mis tähendab, et LSTM mudel ennustab sademeid tihemini kui neid tegelikkuses esineb. See näitab, et mudel prognoosib saju tegelikust pikemat kestvust.

Paljudes olukordades on sademete ennustamata jätmine (VN) rohkem kahju, ohtu või ebamugavust tekitavad kui valehäired. Nendes kontekstides võib POD väärtust olulisemaks pidada kui halba FAR väärtust.

Mudeli ETS oskusskoor on igal ajasammul püsivusprognoosist natuke kõrgem. See näitab, et mudel lisab vähest kuid stabiilset väärtust.

4.3 Võrdlus uuringutega

4.3.1 PredRNN++ kasutades radaripilte

Uuringus [28] on kasutatud videoprognosi süvaõppemudelit PredRNN++ (*Predictive Recurrent Neural Network* ehk ennustav korduv närvivõrk) ning kahe polarisatsiooniga S-kanali (S-band) Doppler radari piltide seeriat, et ennustada tuleviku piltide seeriat ühe tunni ulatuses. Uuringus sooviti luua mudel, mis toetaks teisi lühiprognosi meetodeid. Kasutati andmeid 2015 märtsist 2019 lõpuni ainult ühtelt madalalt mõõtekõrguselt ($0,5^\circ$ *Plan Position Indicator* - PPI² ehk kiire tasapinna nurga indikaator), kusjuures iga pildi vahel oli 15 minutit. Prognoositi 10 pildi seeria põhjal 10 tuleviku pilti (150 minutit), valimid olid moodustatud libiseva aknaga. Lisa 90 minutit ennustati mudeli prognoosi halvenemise visualiseerimiseks. PredRNN++ mudel treeniti superarvutil, ning võttis umbes 240 tundi. Klasside tasakaalustamist ei kasutatud.

Mudeli hindamiseks kasutati kõiki käesolevas töös kasutatud meetrikaid: POD, FAR ja ETS. Skooride saamiseks kasutati igal pikslil teatud läve väärtust. Kõigepealt võrreldi PredRNN++ tulemusi tegelike radaripiltidega (tabel 6), läve väärtus oli mõlemal juhul madal 1dBZ. Tabeli 6 ja käesoleva töö LSTM mudeli tulemuste (tabelis 2) halvenemiste

² PPI – mõõtmistulemused, mis on saadud maapinnale projitseeritud ühe fikseeritud radarikiire tõusunurgal [2].

võrdlused on toodud tabelis 7, kus on arvatud iga meetrika halvenemise protsent LSTM mudeli jaoks ühe tunni ning PredRNN++ jaoks 45 minuti möödudes.

Tabelist 7 on näha, et isegi kui PredRNN++ mudeli FAR halveneb aeglasemalt, langeb selle üldine prognoosioskus (ETS) lühema aja jooksul märgatavalt kiiremini kui LSTM mudelil. POD halvenemised on sarnasema kiirusega, kuid LSTM mudel säilitab ka seda natuke paremini.

Tabel 6: Uuringu [28] (tabel 2) PredRNN++ mudeli meetrikate tulemused

	15 min	30 min	45 min	60 min	Parim tulemus
POD	0,90	0,86	0,83	0,79	1
FAR	0,45	0,53	0,59	0,62	0
ETS	0,46	0,37	0,31	0,21	1

Tabel 7: Uuringu [28] PredRNN++ mudeli ja LSTM mudeli meetrikate halvenemised (%)

	POD	FAR	ETS	Ajaline samm
LSTM	11,4	53,3	28,6	1 h → 2 h
PredRNN++	12,2	37,8	54,3	15 min → 45 min

4.3.2 PredRNN ning ConvLSTM kasutades kohalikke sademete andmeid

Uuringus [9] on kasutatud ConvLSTM (*Convolutional Long Short-Term Memory* ehk konvolutsiooniline pikk lühimälu) ja PredRNN mudeleid, mis treeniti 1835 automatiseeritud ilmajaama 1-minutilise lahutusvõimega sademete (mm) andmete peal. Puuduvad andmed asendati kasutades ajaliselt ja ruumiliselt lähimate naabrite andmeid. Uuringu eesmärk oli luua 6 minutilise ajaloo põhjal 2 tunni ulatuses väga täpne mudel. Andmed olid aastatest 2013 kuni 2021 ning ilmajaamade andmed interpoleeriti ühtlasele ruudustikule. Klasside tasakaalustamist ei tehtud, kuid prooviti kahte erinevat andmestikku: vaid tugevaid sademeid sisaldav andmestik ning algne puhastatud andmestik, millega treenitud mudeleid kasutatakse käesolevas töös võrdluseks.

Uurimuses kasutati hindamiseks peamiselt ETS meetrikat, mis leiti sarnaselt [28] uurimusele üle ruudustiku pinna. Võrdluseks prooviti viit erinevat läve väärtust sademete klassifitseerimiseks, ning nelja erineva perioodi pikkusega ajaloo andmeid. Käesolevas töös valiti võrdlemiseks uurimuse kõige pikema treeningperioodi (8 aastat) parimad tulemused – need vastasid kõige madalamale läve väärtusele 0,1 mm. Tabelis

8 on võrdluseks toodud kõigi võrreldavate mudelite ETS meetrika väärtuste halvenemised prognoosiaja pikenemisel.

On näha sarnaseid tulemusi eelmise alapeatükiga: PredRNN tüüpi mudeli ETS väärtus langeb palju kiiremini, kui lihtsamate LSTM tüüpi mudelitel. Kõige vähem langes ConvLSTM ETS väärtus.

Tabel 8: LSTM mudeli ning Uuringu [9] (tabel 2) mudelite ETS väärtused ja halvenemine

	1. tund	2. tund	Halvenemine (%)
ConvLSTM	0,33	0,25	24,2
PredRNN	0,49	0,21	57,1
LSTM	0,49	0,35	28,6

5 Arutelu

Tulemused näitavad eelkõige LSTM mudeli stabiilsust prognoositava ajahorisondi pikenedes:

- see lisas püsivusprognoosile väikest, kuid järjepidevat väärtust;
- selle POD ja ETS väärtused langesid aeglasemalt kui PredRNN++ mudeli omad uurimuses [28]. LSTM mudelil säilitab sademete tundlikkuse kauem;
- uurimuses [9] oli LSTM mudel taaskord ETS meetrika puhul PredRNN tüüpi mudelist järjekindlam. Alla jäi LSTM mudel oma sugulasele ConvLSTM mudelile, mis toetab järeldust, et lihtsam LSTM tüüpi mudel võib olla ka suurema müraga andmete, nagu uuringu [9] suurema ajalise ja ruumilise lahutusvõimega andmed, puhul täpsuse püsivuses parem.

Tulemuste põhjal saab järeldada, et keerukamad ja suurema ajalise lahutusvõimega mudelid ei taga tingimata stabiilsemat prognoosi, vaid ajahorisondi pikenedes võib täpsus hoopis kiiremini langeda [28][9]. Käesolevas töös kasutatud lihtne LSTM lähenemine võib aidata usaldusväärsemalt ületada lõhe lähimate tundide vahel, mida traditsioonilised numbrilised mudelid piisavalt hästi veel prognoosida ei suuda [3]. Rohkemate eksperimentideta ei ole võimalik kindlalt öelda, mis kõige rohkem tulemusi mõjutas, sest võrdlemiseks kasutatud uurimustes oli käesoleva tööga palju erinevaid aspekte: ajaline lahutus, tasapinnaline või punktipõhine ennustus, kasutatud andmed ning mudelid. Seetõttu piirduti ka vaid halvenemise analüüsimisega (tabelid 7 ja 8).

Ideed lokaalse LSTM mudeli parandamiseks või edasiseks eksperimenteerimiseks on suuresti piiratud saadavalolevate füüsikaliste suuruste ning nende mõõtmis-sagedustega. Esimeseks mõtteks on proovida suuremat ajalist lahutust, kuid keskkonnaportaali avaandmetes on praegu 1 tunnist väiksema sagedusega vaid 10 minuti tuule ning sademete andmed [29]. See lisaks tõenäoliselt mudelile müra, kuid tuleks analüüsida, kas saadud kasu on suurem.

Teine idee on lisada õhupuhtuse parameeter, sest sademe tilgad moodustuvad väikeste tuumade ümber, mis on sattunud pilve [7]. Seda parameetrit Eesti ilmajaamades ei mõõdeta.

Kolmandaks tuleks teha dimensionaalsuse vähendamist ehk eemaldada mitte- või väheolulised füüsikalised suurused. Selleks saab kasutada näiteks permutatsiooni või eemaldamise meetodit: permutatsiooni meetodil segatakse ükshaaval iga parameeter ära, eemaldamise meetodil eemaldatakse kordamööda parameetreid; peale iga eemaldust või segamist mudel treenitakse ning pärast hinnatakse kui palju iga füüsikaline suurus oskust lisas [30][31].

Kasutatud kirjandus

- [1] „Sõnastik“ Copernicus. Accessed: Jan. 05, 2026. [Online.] Available: <https://marine.copernicus.eu/et/explainers/glossary>
- [2] „Kaugseire“ Keskkonnaagentuur. Accessed: Jan 05, 2026. [Online.] Available: <https://www.ilmateenistus.ee/ilmatarkus/mootetehnika/kaugseire/radarid/>
- [3] F. Fabry, Radar meteorology: principles and practice. Cambridge: Cambridge University Press, 2017. pp. 166-169.
- [4] S. Ravuri et al., “Skilful precipitation nowcasting using deep generative models of radar,” *Nature*, vol. 597, pp. 672-677, Sept 2021.
- [5] B. Raj, S. Sahoo, N. Puviarasan and V. Chandrasekar, “Operational Assessment of High Resolution Weather Radar Based Precipitation Nowcasting System,” *Atmosphere*, vol. 15, no. 2:154, Jan 2024, doi: <https://doi.org/10.3390/atmos15020154>
- [6] E. Aguado, J. E. Burt, Understanding weather and climate, 7th ed. USA: Pearson Education, Inc., 2015, pp. 175.
- [7] C. D. Ahrens, R. Henson, *Meteorology Today: An Introduction to Weather, Climate, and the Environment*, 12th ed. Boston, MA, USA: Cengage Learning, Inc., 2017, pp. 168, 357-367.
- [8] R. Prudden et al., “A review of radar-based nowcasting of precipitation and applicable machine learning techniques,” London, England, University of Exeter, White Paper, 2020. doi: <https://doi.org/10.48550/arXiv.2005.04988>
- [9] D. Kong et al., “Precipitation Nowcasting Based on Deep Learning over Guizhou, China”, *Atmosphere*, vol. 14, no.5: 807, Apr. 2020, doi: <https://doi.org/10.3390/atmos14050807>.
- [10] Q. Zheng, Q. Liu, P. Lao and Z. Lu, “Advances in Deep-Learning-based Precipitation Nowcasting Techniques,” *Journal of Tropical Meteorology*, 2024. [Online.] Available: <https://api.semanticscholar.org/CorpusID:272663427>
- [11] Eesti Keskkonnaagentuur, „Avaandmed“ Keskkonnaportaali. . Accessed: Jul. 1, 2024. [Online.] Available: <https://keskkonnaportal.ee/et/avaandmed>.
- [12] J. Hüds, “Pilvede Aluste Kõrguste Analüüs Seilomeetri Andmete Põhjal,” M.S. thesis, School of Science, Tallinn University of Technology, Tallinn, Estonia, 2019.
- [13] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online.] Available: <https://jmlr.csail.mit.edu/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- [14] W. McKinney, “Data structures for statistical computing in Python,” in *Proc. 9th Python in Science Conf.*, 2010, vol. 445, pp. 56–61. [Online.] Available: <https://pub.curvenote.com/01908378-3686-7168-a380-d82bbf21c799/public/mckinney-57fc0d4e8a08cd7f26a4b8bf468a71f4.pdf>

- [15] C. R. Harris et al., “Array programming with NumPy,” *Nature*, vol. 585, pp. 357–362, 2020. doi: 10.1038/s41586-020-2649-2.
- [16] M. Abadi et al., “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2016. [Online]. Available: <https://www.tensorflow.org/>
- [17] F. Chollet et al., “Keras,” 2015. [Online]. Available: <https://keras.io/>
- [18] M. L. Waskom, “Seaborn: Statistical data visualization,” *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021. doi: <https://doi.org/10.21105/joss.03021>.
- [19] J. D. Hunter, “Matplotlib: A 2D graphics environment,” *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [20] OpenAI, “ChatGPT,” 2024. [Online]. Available: <https://chat.openai.com/>
- [21] Microsoft, “Microsoft Copilot,” 2024. [Online]. Available: <https://copilot.microsoft.com/>
- [22] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” in *Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining (KDD)*, 2019, pp. 2623–2631.
- [23] J. Brownlee, “How to Reduce Overfitting With Dropout Regularization in Keras,” *Machine Learning Mastery*, 2020. Accessed: Jan. 06, 2026. [Online.] Available: <https://machinelearningmastery.com/how-to-reduce-overfitting-with-dropout-regularization-in-keras/>
- [24] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection,” 2018. [Online.] Available: <https://arxiv.org/pdf/1708.02002>
- [25] “Forecast Verification methods Across Time and Space Scales.” *cawcr.gov.au*, Accessed: Dec. 28, 2025. [Online.] Available: <https://www.cawcr.gov.au/projects/verification/>
- [26] “Application and Verification ECMWF products in Member States and Co-operating States,” Report, 2006. [Online.] Available: https://www.ecmwf.int/sites/default/files/elibrary/2006/74335-application-and-verification-ecmwf-products-member-states-and-co-operating-states_0.pdf.
- [27] J. Brownlee, “A Gentle Introduction to Threshold-Moving for Imbalanced Classification,” Jan. 5, 2021. [Online]. Available: <https://machinelearningmastery.com/threshold-moving-for-imbalanced-classification/>
- [28] S. M. Bonnet, A. Evsukoff, C. A. Morales Rodriguez, “Precipitation Nowcasting with Weather Radar Images and Deep Learning in São Paulo, Brasil,” *Atmosphere*, vol. 11, no. 11: 1157, Oct. 2020, doi: 10.3390/atmos11111157.
- [29] „Kliimaandmestiku kirjeldus” *Keskkonnaportaals*. Accessed: Jan. 06, 2026. [Online.] Available:

<https://keskkonnaportaal.ee/et/avaandmed/kliimaandmestik/kliimaandmestiku-kirjeldus>

- [30] scikit-learn developers, „Permutation feature importance“ scikit-learn, Accessed: Jan. 05, 2026. [Online.] Available: https://scikit-learn.org/stable/modules/permutation_importance
- [31] J. Drams, „Ablation Studies“ ml.recipes.com. Accessed: Jan 05, 2026. [Online.] Available: <https://ml.recipes/notebooks/6-ablation-study>

Lisad

Lisa 1 – Optimeeritud modeli struktuur

```
best_model = Sequential()
best_model.add(Input(shape=(n_timesteps, n_features)))

best_model.add(
    LSTM(
        43,
        return_sequences=True,
        dropout=0.07028726141306774,
        recurrent_dropout=0.0515254853751841
    )
)

best_model.add(
    LSTM(
        58,
        return_sequences=False,
        dropout=0.07028726141306774,
        recurrent_dropout=0.0515254853751841
    )
)

best_model.add(Dense(3, activation='sigmoid'))

best_model.compile(optimizer=Adam(learning_rate=0.009150114118798836,
                                loss=keras.losses.BinaryFocalCrossentropy(
                                    apply_class_balancing=False,
                                    alpha=0.773732098296195,
                                    gamma=2.716007422792222)
                                )

checkpoint = ModelCheckpoint(
    # PATH,
    monitor='val_loss',
    save_best_only=True,
    mode='min',
    verbose=1
)

early_stop = EarlyStopping(monitor='val_loss', patience=30,
                           min_delta=1e-5, restore_best_weights=True,
                           mode='min')

result = best_model.fit(X_train, y_train,
                        validation_data=(X_val, y_val),
                        batch_size=32,
                        epochs=100,
                        shuffle=False,
                        callbacks=[early_stop, checkpoint],
                        verbose=2)
```

Lisa 2 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Marili Arus

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Sademete lühiajaline prognoos masinõppe abil“, mille juhendaja on Sander Rikka,

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.

06.01.2026