

Ataskaitinė informatikos inžinerijos krypties
doktorantų konferencija

Ataskaita už 2017 – 2018 mokslo metus

Doktorantas: Julius Venskus

Vilnius, 2018

Bendra informacija

- **Disertacijos pavadinimas:** „Giliojo mašininio mokymo taikymas jūrų transporto eismo duomenims klasifikuoti“
- **Darbo vadovas:** dr. Povilas Treigys
- **Konsultantas:** dr. Arūnas Andziulis
- **Doktorantūros pradžia:** 2016 m.
- **Planuojama doktorantūros pabaiga:** 2020 m.

Informacija apie tyrimą

■ Tyrimo objektas:

- jūrų transporto priemonių eismo duomenys

■ Tyrimo tikslas:

- ištirti ir pasiūlyti būdus jūrų transporto duomenų klasifikavimui ir neįprasto eismo aptikimui.

■ Tyrimo uždaviniai:

- apžvelgti esamus jūrų transporto klasifikavimo ir neįprasto eismo aptikimo būdus AIS (automatinės identifikavimo sistemos) duomenyse;
- ištirti klasikinius jūrų transporto klasifikavimo ir neįprasto eismo aptikimo AIS duomenyse būdus;
- ištirti ir pritaikyti giliuoju mokymusi grindžiamus algoritmus transporto duomenims klasifikuoti;
- atlikti eksperimentinį klasifikatorių vertinimą.

■ Planuojami rezultatai:

- jūrų transporto eismo klasifikavimo ir neįprasto eismo aptikimo algoritmus įgyvendinantis programų sistemos prototipas.

2017/2018 darbo planas

■ Studijų planas:

- egzaminas „**Loginiai dirbtinio intelekto pagrindai**“, 7 ECTS, prof. habil. dr. Albertas Čaplinskas

■ Mokslinių tyrimų publikavimas:

- analitinio tyrimo rezultatų publikavimas recenzuojamame periodiniame moksliniame leidinyje.

■ Dalyvavimas konferencijose, seminaruose, kitose doktorantų mobilumo veiklose:

- dalyvavimas tarptautinėje arba nacionalinėje konferencijoje pristatant disertacijos teorinio tyrimo rezultatus.
- sudalyvauti bent vienoje tarptautinėje doktorantų mokykloje.

Ataskaita už 2017/2018 mokslo metus (1)

■ Išlaikyti egzaminai:

- egzaminas „**Loginiai dirbtinio intelekto pagrindai**“, 7 ECTS,
Vertinimo komisija: prof. habil. dr. Albertas Čaplinskas, prof. dr. Saulius Gudas, doc. dr. Audronė Lupeikienė – **Įvertinimas 9 (l. gerai)**

■ Mokslinių tyrimų publikavimas:

- Venskų, Julius; Treigys, Povilas; Bernatavičienė, Jolita; Tamulevičius, Gintautas; Medvedev, Viktor;

Unsupervised Neural Network Retraining: Real-time Maritime Traffic Anomaly Detection Based on History Data Embedding,
Neural network world. Czech Technical University in Prague. ISSN 1210-0552. - **Įteiktas**

■ Pranešimai konferencijose:

- Venskų, Julius; Treigys, Povilas; Bernatavičienė, Jolita; **Retraining strategies of modified SOM for abnormal marine traffic detection**;
Materials, Methods & Technologies 2018 : 20th International conference. Elenite, Bulgaria, June 26-30, 2018. International Scientific Events

Ataskaita už 2017/2018 mokslo metus (2)

■ Dalyvavimas seminaruose, kitose doktorantų mobilumo veiklose:

- **BigSkyEarth Training School 2018:** GPU-based analytics and data science. COST Action: TD1403, Training School Title: Third training school, Venue: Centro De Tecnologías De Interacción Visual Y Comunicaciones, Vicomtech, San Sebastian, Spain, Training School Dates: 2018-04-03 09:00:00 - 2018-04-09 18:00:00
- Venskų, Julius; Treigys, Povilas; Bernatavičienė, Jolita; Medvedev, Viktor. **Retraining strategies for abnormal marine traffic detection //** 9th International workshop on Data Analysis Methods for Software Systems (DAMSS), Druskininkai, Lithuania, November 30 - December 2, 2017. Vilnius : Vilniaus universitetas, 2017. ISBN 9789986680642. p. 54.
- Pranešimas: ***VU Duomenų mokslo ir skaitmeninių technologijų instituto KSG seminare:*** Venskų Julius, Savi-organizuojančio neuroninio tinklo (SOM) ir virtualaus feromono integravimas neįprastam laivų eismo aptikimui, 2018m. Vilnius

Ataskaitinė informatikos inžinerijos krypties
doktorantų konferencija

Mokslinė ataskaita

Doktorantas: Julius Venskus

Vadovas: dr. Povilas Treigys

Konsultantas: prof. dr. Arūnas Andziulis

Vilnius, 2018

Įvadas(2)

Jūrinio sektoriaus anomalija

- E. Martineau ir J. Roy „Maritime Anomaly Detection: Domain Introduction and Review of Selected Literature“
 - Praeities duomenyse atrastų **šablonų atpažinimas** ir naujai gaunamų duomenų klasifikavimas pagal surastus šablonus. Jei nauji duomenys **neatitinka surastų šablonų** (normaliam modeliui) jie klasifikuojami kaip anomalūs (nejprasti).
- Nagrinėjama jūrinio sektoriaus anomalijos dalis – nejprastas laivų eismas.
 - Laivo geografinė padėtis.
 - Laivo padėtis kitų laivų ar kliūčių atžvilgiu.
 - Laivo greitis ir kryptis.
 - Aukščiau išvardintų savybių kitimas laiko atžvilgiu.
 - Kitų sąlygų įtaka aukščiau išvardintoms savybėms.
- Dažniausiai aptinkami straipsniuose terminai “outliers,” “anomalies,” “unusual,” „abnormal“, „marine traffic“.

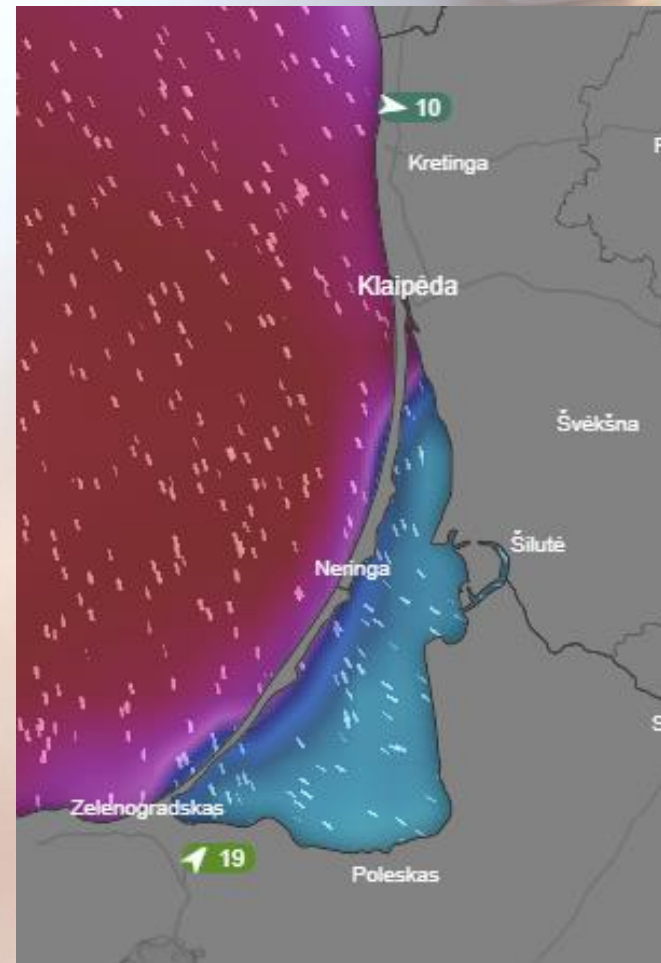
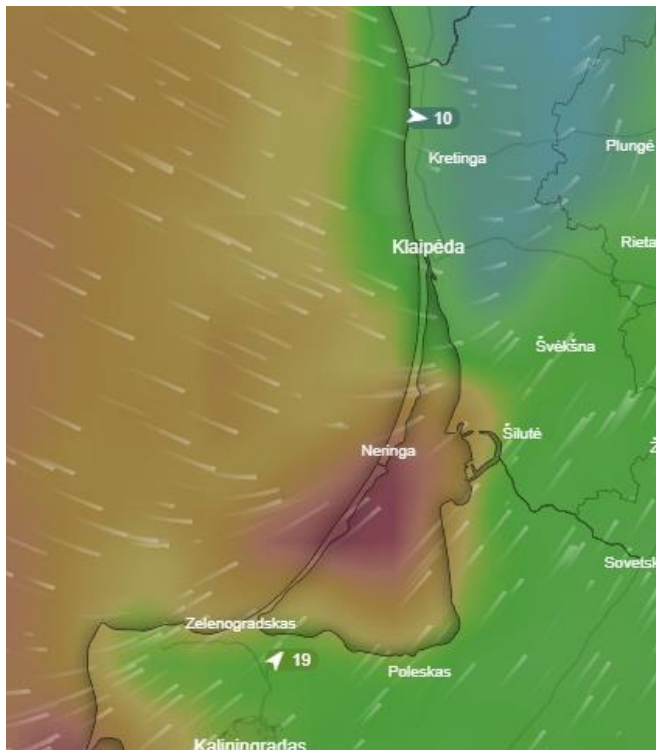
Automatinė indentifikavimo sistema (AIS)

Automatinė indentifikavimo sistema (AIS)



Duomenų šaltiniai (1)

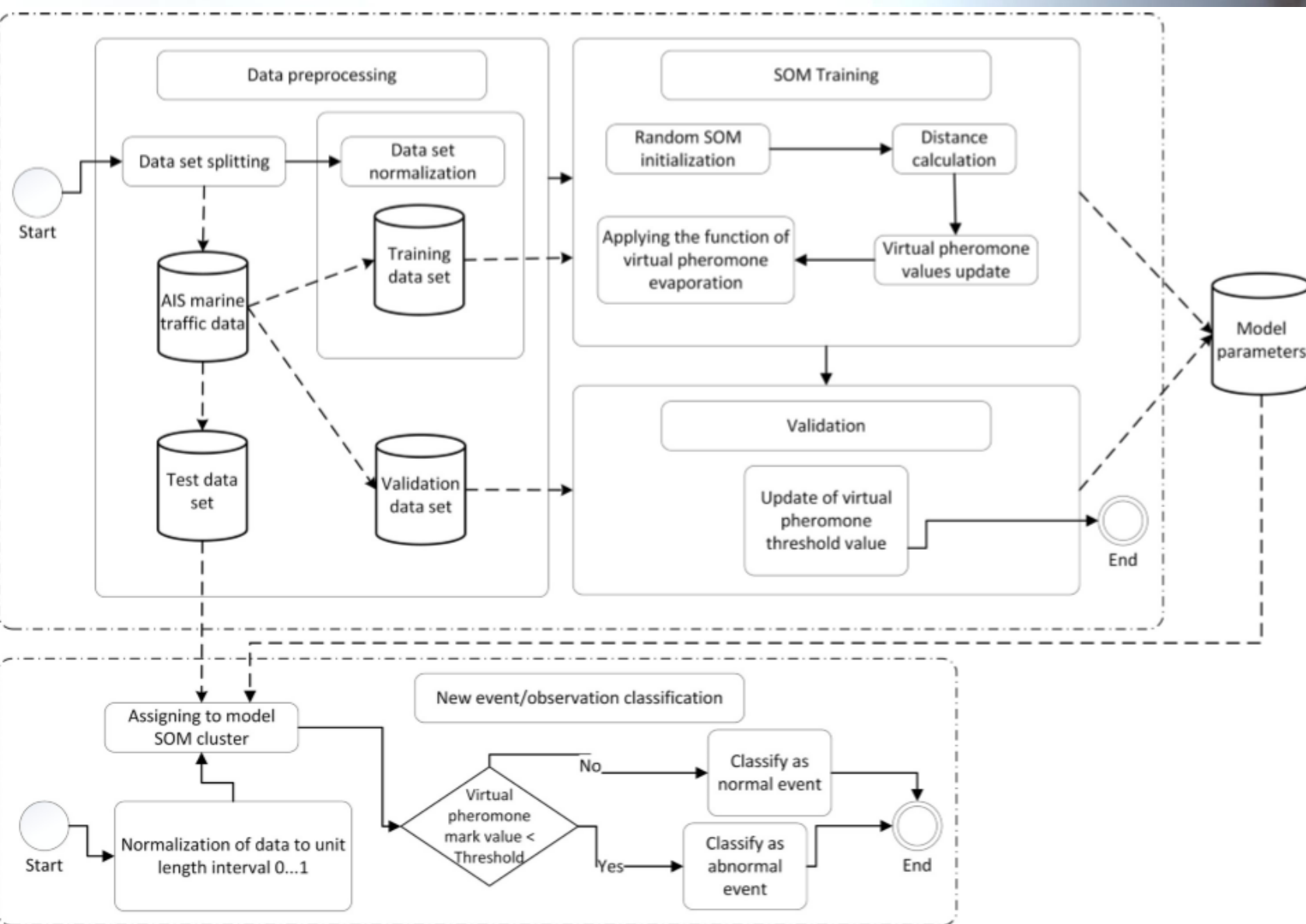
- Meteorologiniai duomenys
 - Bangavimo kryptis ir stiprumas
 - Vėjo kryptis ir stiprumas
- Šaltinio tipai
 - Tiesioginiai matavimai (LHMT)
 - Modelio ECMWF (European Center for Medium range Weather Forecasting)





2016/2017 Tyrimas

2016/2017 sukurtas modifikuotas SOM skirtas aptikti neįprastus laivų eismo atvejus.



2016/2017 Rezultatai

Table 4. Influence of the neighbourhood function on the classification accuracy when the SOM grid dimension is 20x20

	Neighbourhood function	TP	FP	TN	FN	Precision	Sensitivity
Expert		1681	0	27167	0	1	1
SOM_GMM	Gaussian	1489	81	27086	192	0.948	0.886
	Gaussian	1477	68	27099	204	0.956	0.879
	Triangular	1241	122	27045	440	0.911	0.738
SOM_Pheromone	Bubble	1454	68	27099	227	0.955	0.865
	Cut Gaussian	1479	65	27102	202	0.958	0.880
	Mexican hat	1509	51	27116	172	0.967	0.898

Table 5. Influence of the neighbourhood function on the classification accuracy when the SOM grid dimension is 25x25

	Neighbourhood function	TP	FP	TN	FN	Precision	Sensitivity
Expert		1681	0	27167	0	1	1
SOM_GMM	Gaussian	1495	80	27087	186	0.949	0.889
	Gaussian	1491	59	27108	190	0.962	0.887
	Triangular	1288	117	25948	1495	0.917	0.463
SOM_Pheromone	Bubble	1455	63	27104	226	0.955	0.865
	Cut Gaussian	1498	55	27112	183	0.958	0.866
	Mexican hat	1512	50	27117	169	0.968	0.899

2016/2017 Rezultatai

Table 6. Influence of the SOM grid dimension on the classification accuracy of SOM_Pheromone and SOM_GMM algorithms

Grid dimension	SOM_Pheromone		SOM_GMM	
	Precision	Sensitivity	Precision	Sensitivity
10x10	0.919	0.773	0.867	0.780
15x15	0.933	0.814	0.921	0.834
20x20	0.967	0.898	0.948	0.886
25x25	0.968	0.899	0.949	0.889
30x30	0.961	0.897	0.948	0.888
35x35	0.948	0.893	0.932	0.877
40x40	0.918	0.886	0.919	0.875

Table 7. Classification results of the Passenger vessels data set (normal states: 8193, abnormal states: 1457)

Method	TP	FP	TN	FN	Precision	Sensitivity
SOM_GMM	1314	17	8176	143	0.987	0.902
SOM_Pheromone	1328	18	8175	123	0.987	0.911

Table 8. Classification results of the Tugs and Pilot vessels data set (normal states: 12298, abnormal states: 1153)

Method	TP	FP	TN	FN	Precision	Sensitivity
SOM_GMM	971	9	12289	182	0.991	0.842
SOM_Pheromone	978	9	12289	175	0.991	0.848

2017/2018 Tyrimas

Unsupervised Neural Network Retraining: Real-time Maritime Traffic Anomaly Detection Based on History Data Embedding.

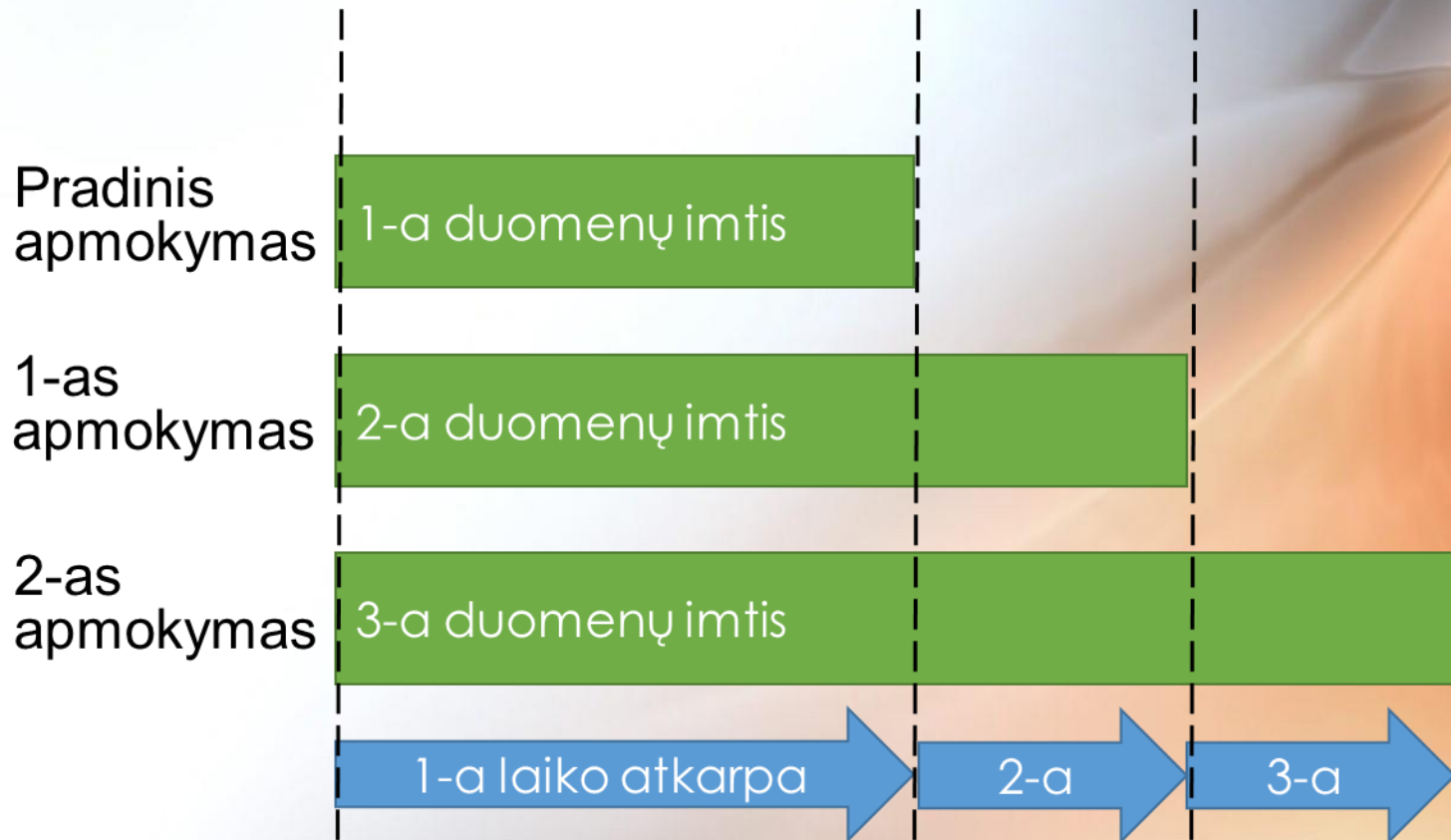
Neprižiūrimasis neuroninio tinklo permokymas: Laivų eismo anomalijos aptikimas realiu laiku paremtas istorinių duomenų įterpimu.

Permokymo strategijos

- Tiriamos permokymo strategijos:
 - Strategija I: Permokymas su visais turimais duomenimis (batch learning).
 - Strategija II: Permokymas tik su naujais duomenimis, naudojant senas neuronų reikšmes.
 - Strategija III: Permokymas su naujais duomenimis ir dalimi istorinių duomenų.

Strategija I

Permokymų seka



Strategija I

Rezultatai

Table 1: Selection of learning rate

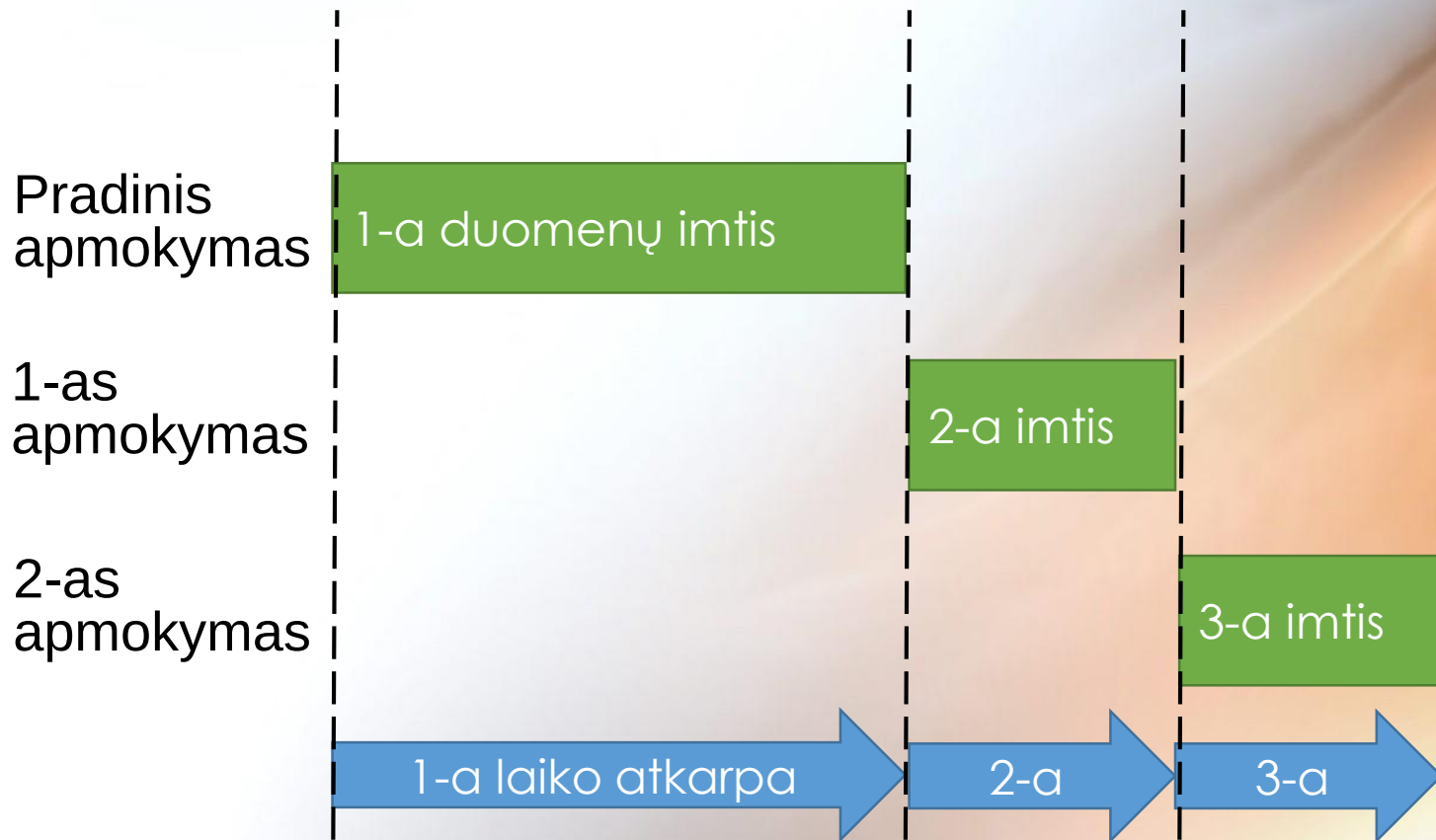
Learning rate	TP	FP	TN	FN	Precision	Sensitivity
0.005	924	519	26648	757	0.6403	0.5497
0.010	943	505	26662	738	0.6512	0.5610
0.015	957	498	26669	724	0.6577	0.5693
0.020	963	487	26680	718	0.6641	0.5729
0.025	968	478	26689	713	0.6694	0.5758
0.030	976	471	26696	705	0.6745	0.5806
0.035	986	468	26699	695	0.6781	0.5866
0.040	998	461	26706	683	0.6840	0.5937
0.050	1025	445	26722	656	0.6973	0.6098
0.060	1066	413	26754	615	0.7208	0.6341
0.070	1109	394	26773	572	0.7379	0.6597
0.100	1197	303	26864	484	0.7980	0.7121
0.200	1431	135	27032	250	0.9138	0.8513
0.300	1486	81	27086	195	0.9483	0.8840
0.400	1500	55	27112	181	0.9646	0.8923
0.500	1510	52	27115	171	0.9667	0.8983
0.600	1507	54	27113	174	0.9654	0.8965
0.700	1502	59	27108	179	0.9622	0.8935

Table 2: Training Strategy I performance at learning rate 0.5

Stage	TP	FP	TN	FN	Precision	Sensitivity
Testing (model error)	1510	52	27115	171	0.9667	0.8983
Testing (general error)	1868	69	33890	233	0.9644	0.8891

Strategija II

Permokymų seka



Strategija II Rezultatai

Table 3: Strategy II performance on model test data

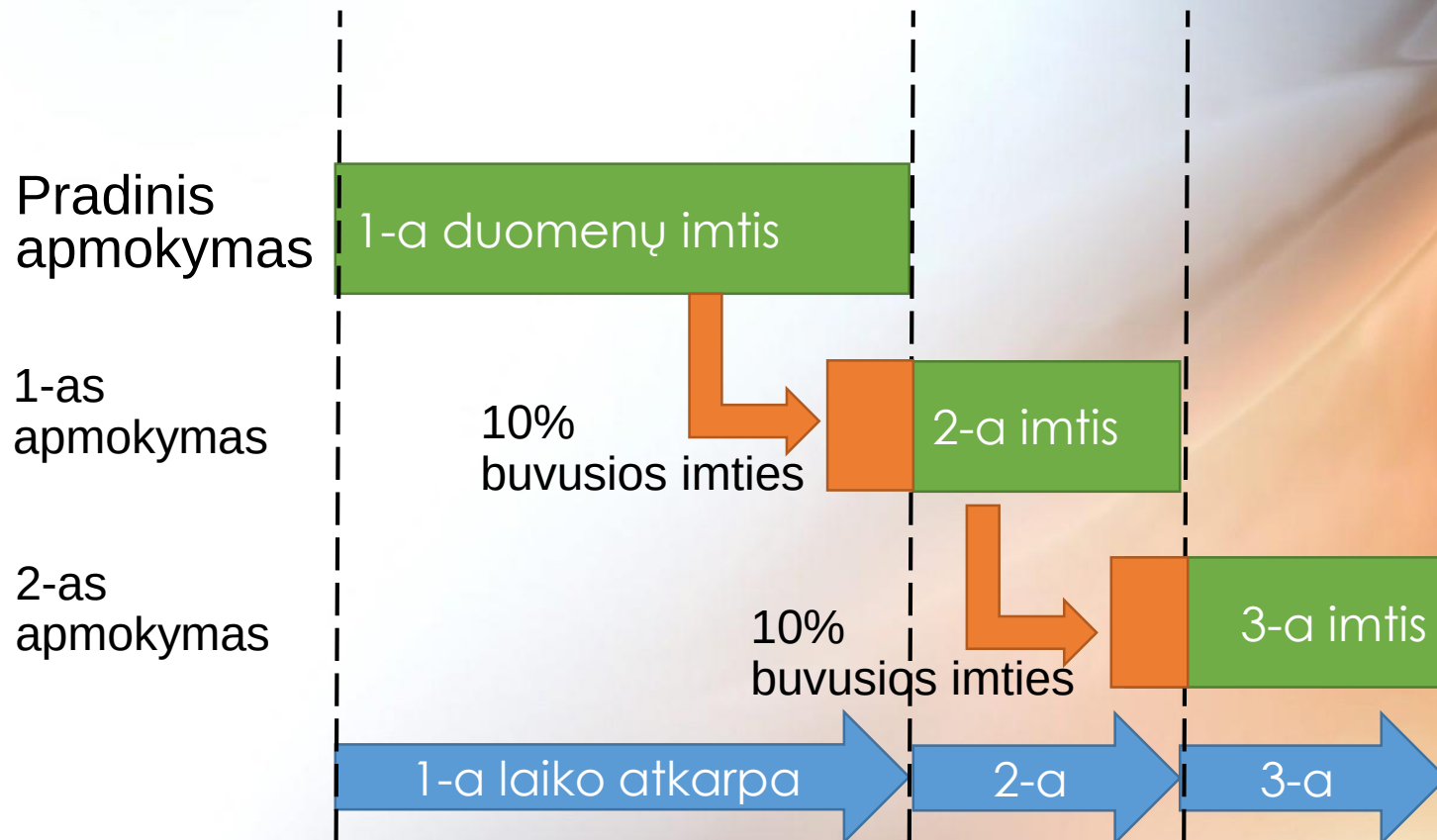
No.	TP	FP	TN	FN	Precision	Sensitivity
1	1364	241	26926	317	0.8498	0.8114
2	1329	280	26887	352	0.8260	0.7906
3	1359	252	26915	322	0.8436	0.8084
4	1364	274	26893	317	0.8327	0.8114
5	1356	253	26914	325	0.8428	0.8067
6	1335	253	26914	346	0.8407	0.7942
7	1314	251	26916	367	0.8396	0.7817
8	1332	258	26909	349	0.8377	0.7924
9	1367	237	26930	314	0.8522	0.8132
10	1338	240	26927	343	0.8497	0.7960
max					0.8522	0.8132
min					0.8260	0.7817
average					0.8413	0.8011
stdev					0.0079	0.0115

Table 4: Retraining Strategy II performance at learning rate 0.025

Stage	TP	FP	TN	FN	Precision	Sensitivity
Testing (model error)	1500	98	27069	181	0.9387	0.8923
Testing (general error)	1836	122	33837	265	0.9377	0.8739

Strategija III

Permokymų seka



Strategija III Rezultatai

Table 6: Retraining Strategy III performance at learning rate 0.003

Stage	TP	FP	TN	FN	Precision	Sensitivity
Testing (model error)	1527	73	27094	154	0.9544	0.9084
Testing (general error)	1866	91	33868	235	0.9535	0.8881

Table 7: Retraining Strategies I-III performance on Cargo data set

Strategy	Precision	Sensitivity	Relative time
Strategy I	0.9644	0.8891	1
Strategy II	0.9377	0.8739	0.4471
Strategy III	0.9535	0.8881	0.6832

Ateities darbai

- Surinkti laivų eismo ir meteorologinius duomenis iš kelių skirtingų geografinių regionų.
- Pritaikyti kelių tipų giliuosius neuroninius tinklus, kad atpažintų neįprasto laivų eismo atvejus, ir ištirti jo klasifikavimo rodiklius bei paliginti su prieš tai tirtais algoritmais.



Ačiū
Klausimai?