



**Vilniaus universitetas**  
**Duomenų mokslo ir skaitmeninių**  
**technologijų institutas**  
**LIETUVA**



**DOKTORANTŪROS METINĖ ATASKAITA**

2021 m. spalio mėn. 1 d. – 2022 m. rugsėjo mėn. 30 d.

**INFORMATIKOS STUDIJŲ PROGRAMOS**  
**DOKTORANTĖ MARTA KARALIUTĖ**

**Disertacijos pavadinimas:** Erdvės-laiko duomenų klasifikavimas naudojant diskriminantines funkcijas

**Vadovas:** prof. dr. Kęstutis Dučinskas

**Konsultantas:** prof. habil. dr. Gintautas Dzemyda

**Doktorantūros laikotarpis:** 2017 – 2022

**Studijų metai:** IV (2021/2022)

► **Tyrimo objektas:**

Erdvės-laiko duomenų klasifikavimo metodai

► **Tyrimo tikslas:**

Atlikti erdvės-laiko duomenų statistinį klasifikavimą naudojant diskriminavimo funkciją bei klasifikavimo klaidų tikimybes. Išvestų formulių pagrindu sukurti algoritmus ir juos pritaikyti realių duomenų analizei.

## **Tyrimo uždaviniai:**

- ▶ Bajeso ir įterptinių Bajeso diskriminantinių funkcijų pritaikymas Erdvės-laiko Gausinių duomenų generatyvinio prižiūravimo klasifikavimo uždaviniuose.
- ▶ Siūlomų klasifikavimų procedūrų tyrimas ir palyginimas, panaudojant klasių žymių skirstinius.

# Lentelė 1. Visų studijų planas ir jo vykdymo suvestinė

Studijų metai	Egzaminai		Dalyvavimas konferencijose		Publikacijos		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė
I (2017/2018)	2	2					
II (2018/2019)	2	2	1	2	1	0	
III (2019 10/2020 08) (2021 09/2021 09)			1		1	1	Publikuota
<b>IV (2021/2022)</b>				<b>1</b>	<b>1 (skola iš II metų)</b>	<b>2</b>	<b>Įteikta</b>
Iš viso:	4	4	2	3	2		

## Lentelė 2. Ataskaitinių metų darbo planas ir jo įvykdymas

### Dalyvavimas konferencijose

Planas	Įvykdyta	Konferencijos tipas
Tyrimo rezultatų pristatymas tarptautinėje mokslinėje konferencijoje	M. Karaliutė, K. Dučinskas, Supervised generic model for classification of Gaussian spatio-temporal areal. Participated in the 31st International Biometric Conference, 10-15 July, 2022, Riga, Latvia	Tarptautinė konferencija

## Lentelė 2. Ataskaitinių metų darbo planas ir jo įvykdymas

Publikacijos			
Planas	Įvykdyta	Būklė	Publikacijos tipas
Žurnalas, turintis cituojamumo rodiklį Clarivate Analytics Web of Science duomenų bazėje	Marta Karaliutė, Kęstutis Dučinskas, Supervised generative classification of areal spatio-temporal Gaussian data, Pattern Recognition Letters, 2022	Įteikta	(impact factor) CA WoS duomenų bazėje
Žurnalas, turintis cituojamumo rodiklį Clarivate Analytics Web of Science duomenų bazėje	Marta Karaliutė, Kęstutis Dučinskas, Performance of the supervised generative classifiers of spatio-temporal areal data using various spatial autocorrelation indexes. Nonlinear analysis: modelling and control. 2022	Įteikta	(impact factor) CA WoS duomenų bazėje

## Lentelė 3. Visų mokslinių tyrimų ir disertacijos rengimo etapai (I)

	Darbo pavadinimas	Atlikimo terminai	Pastabos
1.	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje).	2017 m. spalio – 2018 m. balandis	Parengta mokslinės literatūros disertacijos tema apžvalga.
2.	Mokslinio tyrimo vykdymas: 2.1. Tyrimo metodikos sudarymas:  2.2. Teorinis tyrimas:  2.3. Empirinis tyrimas:	2017 m. spalio  2018 m. gegužė - 2019 m. rugsėjis  2019 m. spalio – 2021 m. rugsėjis	Atlikta.  Atlikta.  Publikacijos (generuotų duomenų ir realių duomenų klasifikavimo uždaviniai).



## Lentelė 3. Visų mokslinių tyrimų ir disertacijos rengimo etapai (II)

	<b>Darbo pavadinimas</b>	<b>Atlikimo terminai</b>	<b>Pastabos</b>
	2.4. Gautų duomenų analizė, apibendrinimas, išvadų parengimas.	2021 m. spalio – 2021 m. gruodis	Generuotų ir realių duomenų klasifikavimo uždavinių rezultatų lyginimas
<b>3.</b>	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:	2022 m. sausis – 2022 m. balandis	Vykdoma.
<b>4.</b>	Daktaro disertacijos parengimas ir svarstymas padalinyje	2022 m. gegužė	Vykdoma.
<b>5.</b>	Daktaro disertacijos gynimas	2022 m. rugsėjis	

Nagrinėjamos šios erdvės-laiko kovariacinės struktūros:

Erdvės koreliacijos matrica:

1. eksponentiniu semivariogramos modeliu,

t. y.  $R = (r_{ij})$ , kur  $r_{ij} = r(|s_i - s_j|) = e^{-|s_i - s_j|/\varphi}$ .

2.  $R = (I_n + \alpha H)^{-1}$ , kur  $H = D - W$ ,

$$w_{ij} = \begin{cases} 0, & \text{kai } i = j \\ 1, & \text{kai } s_j \in N_i \\ 0, & \text{kitais atvejais} \end{cases}, \quad d_{ij} = \begin{cases} \sum_{s_j \in N_i} w_{ij}, & \text{kai } i = j \\ 0, & \text{kitais atvejais} \end{cases}$$

Laiko kovariacijos matrica  $\Sigma$  skaičiuojama pagal Yule-Walker lygtis AR(1) modeliui.

Apriorinės tikimybės taške  $s_i$  laike  $t = m + 1$  yra apibrėžiamos:

1. Laiko svertinis vidurkis (Temporal Weighted Moving Average (TWMA)):

$$\pi_{1t}(s_i, m + 1) = \frac{\sum_{t=1}^m y_i^t t}{(1 + m)m/2}.$$

2. Laiko ir erdvės svertinis vidurkis (Spatial Temporal Weighted Moving Average (STWMA)):

$$\pi_{1ts}(s_i, m + 1) = \frac{\sum_{t=1}^m y_i^t t + \sum_{j:s_j \in N_i} \sum_{t=1}^m y_j^t t}{(1 + m)m/2 (1 + |N_i|)}$$

3. Laiko ir erdvės nesvertinis vidurkis (Spatial Temporal Unweighted Average (STUA)):

$$\pi_{1s}(s_i, m + 1) = \frac{\sum_{t=1}^m (y_i^t + \sum_{j:s_j \in N_i} y_j^t)}{m(1 + |N_i|)}.$$

4. Globalaus Morano I (Global Moran's I) atveju:

$$\pi_{1M}(s_i, m + 1) = \frac{1}{1 + \exp(-I(m)y_i^*(m))}$$

5. Geary's C atveju:

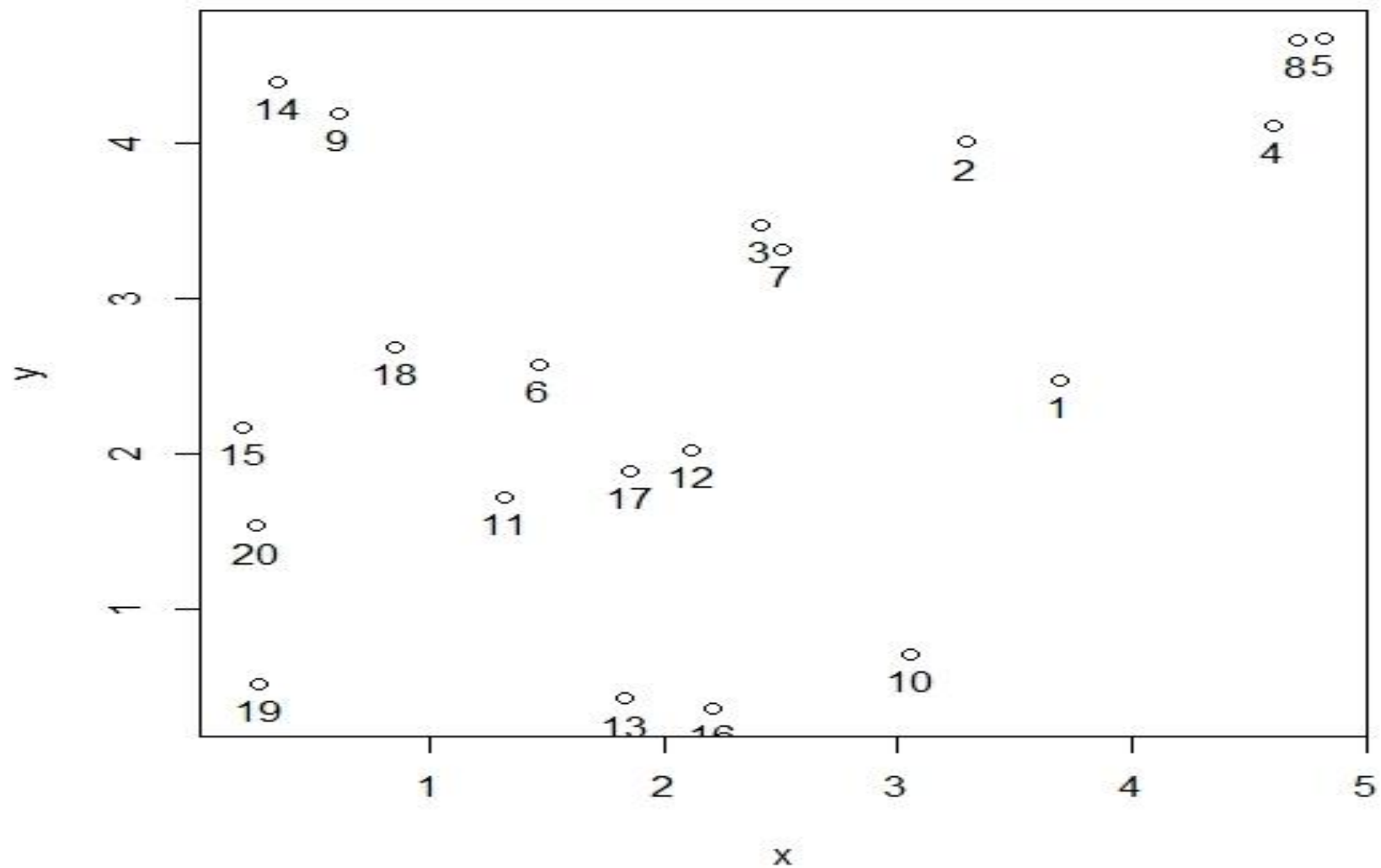
$$\pi_{1C}(s_i, m + 1) = \frac{1}{1 + \exp(-C(m)y_i^*(m))}$$

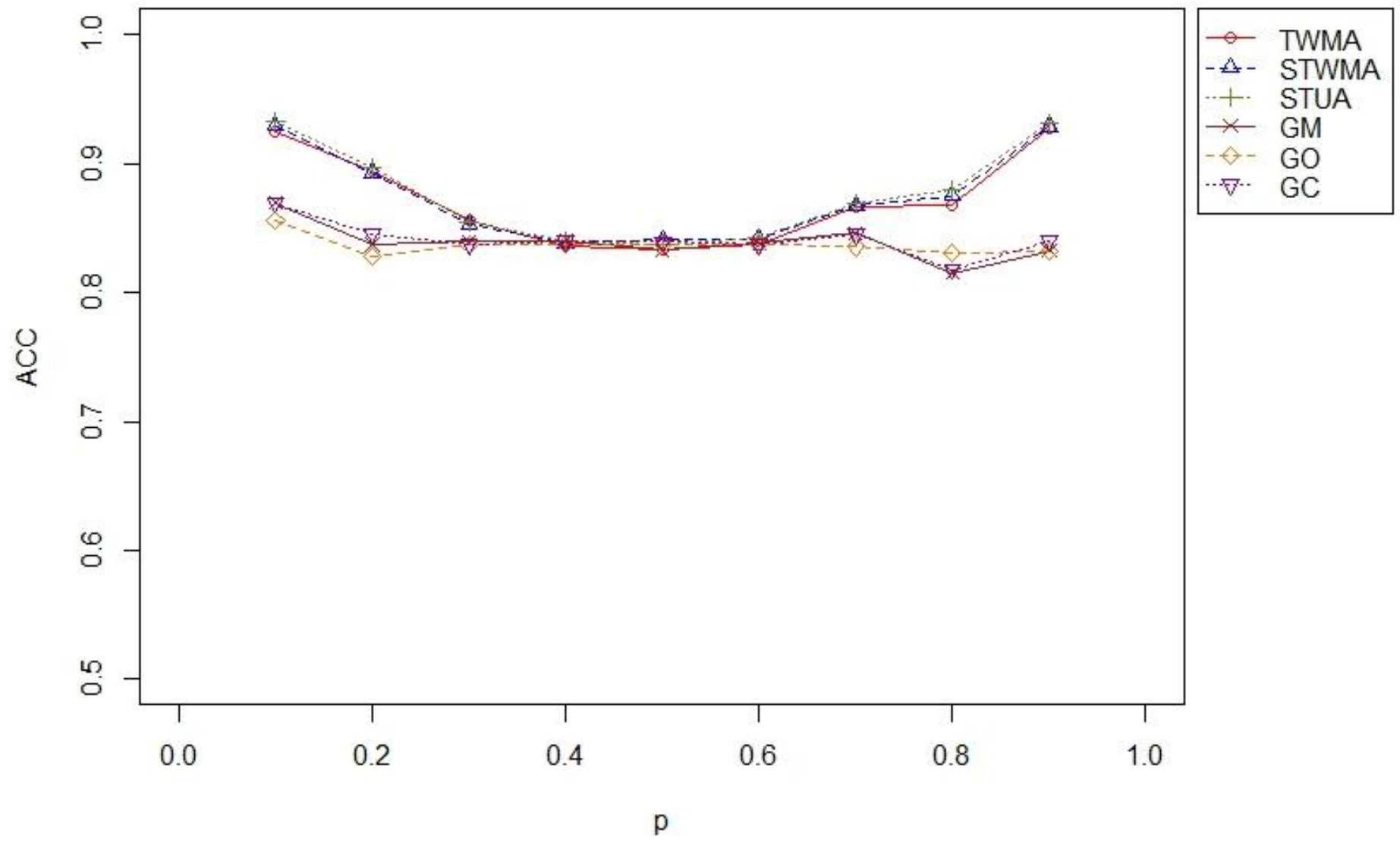
6. Getis-Ord G atveju:

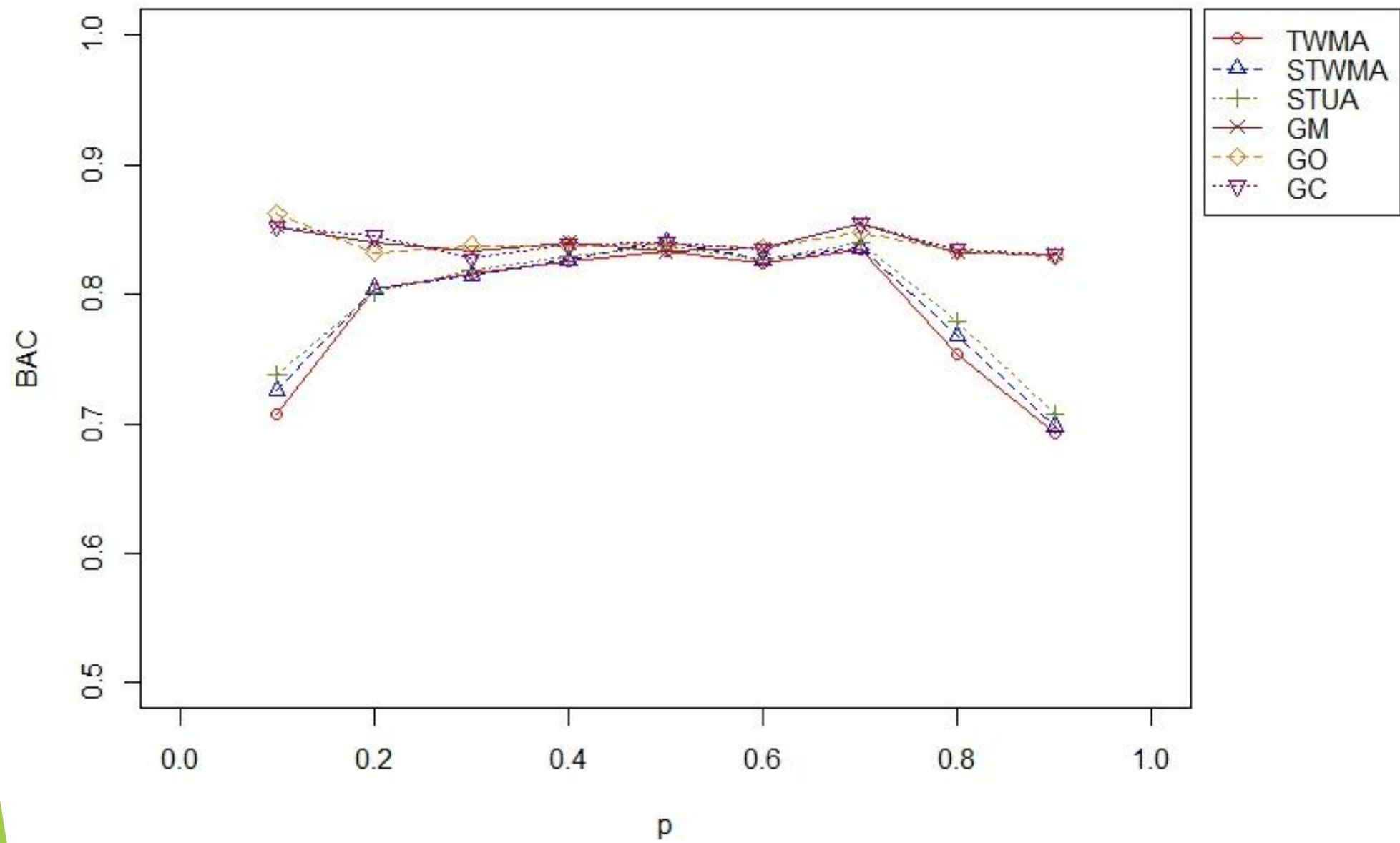
$$\pi_{G1t}(s_i, m + 1) = \frac{1}{1 + \exp(-G(m)y_i^*(m))}$$

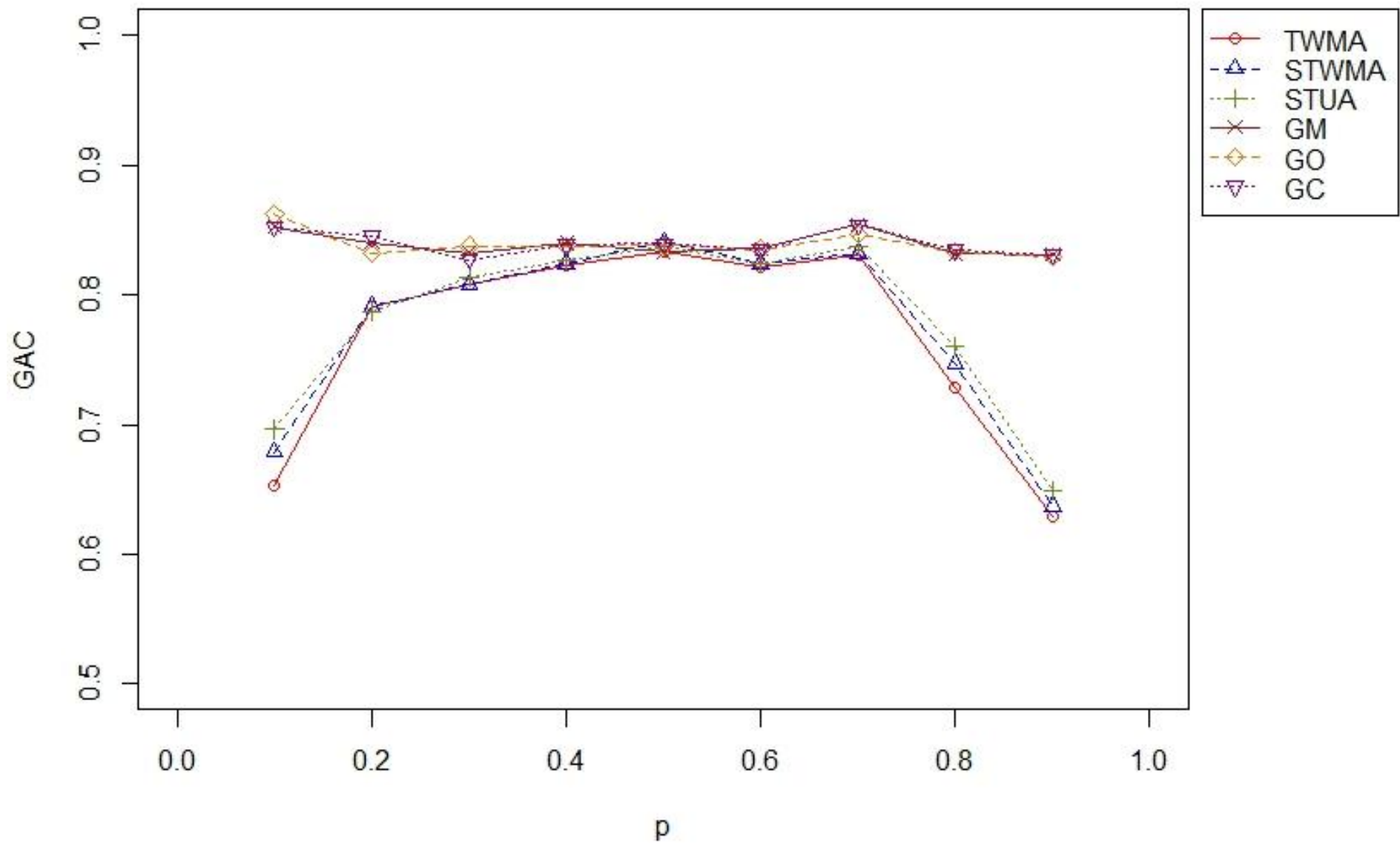
kur  $y_i^*(t) = 2y_i(t) - 1$ .

# Gauti rezultatai (1)





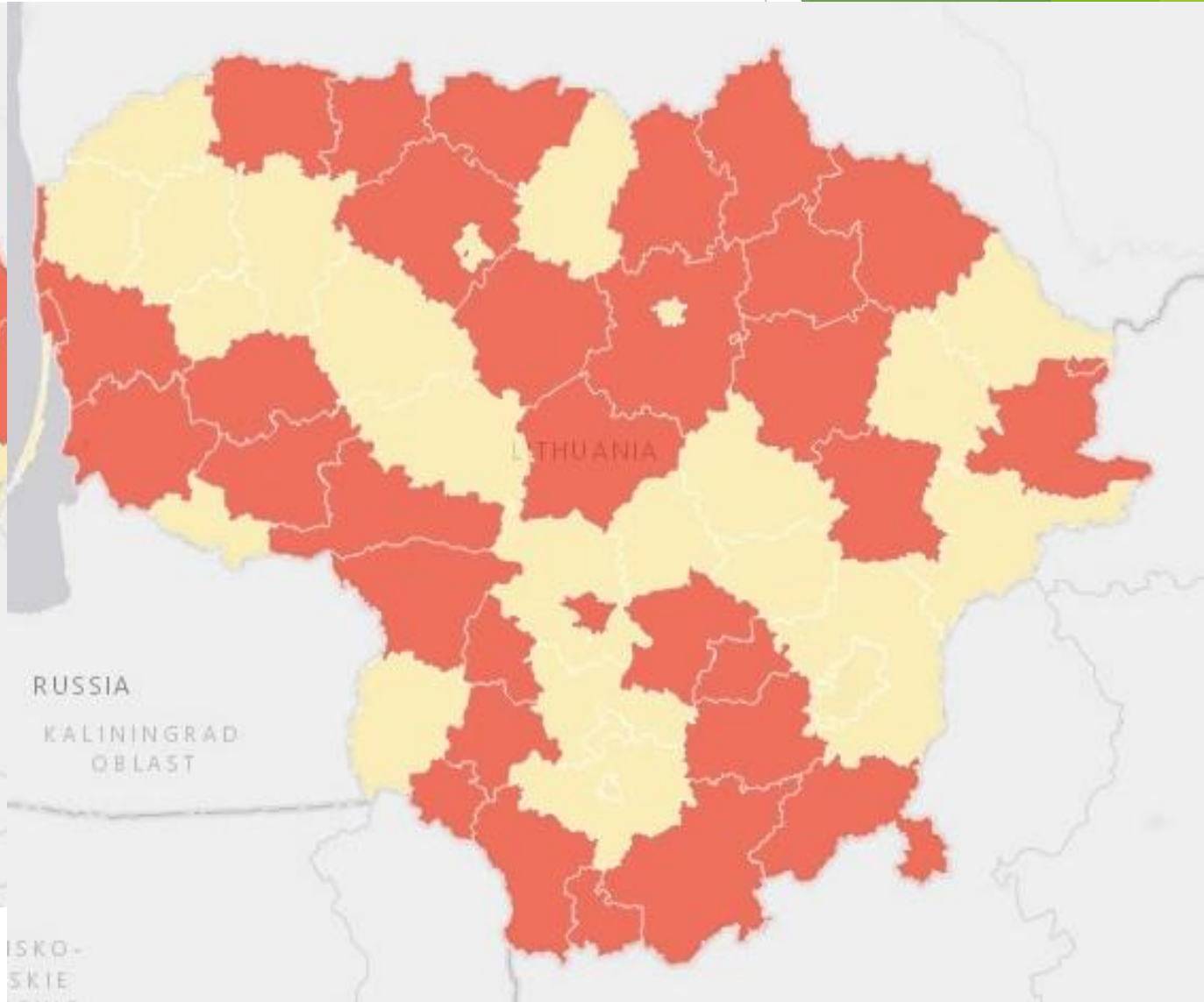
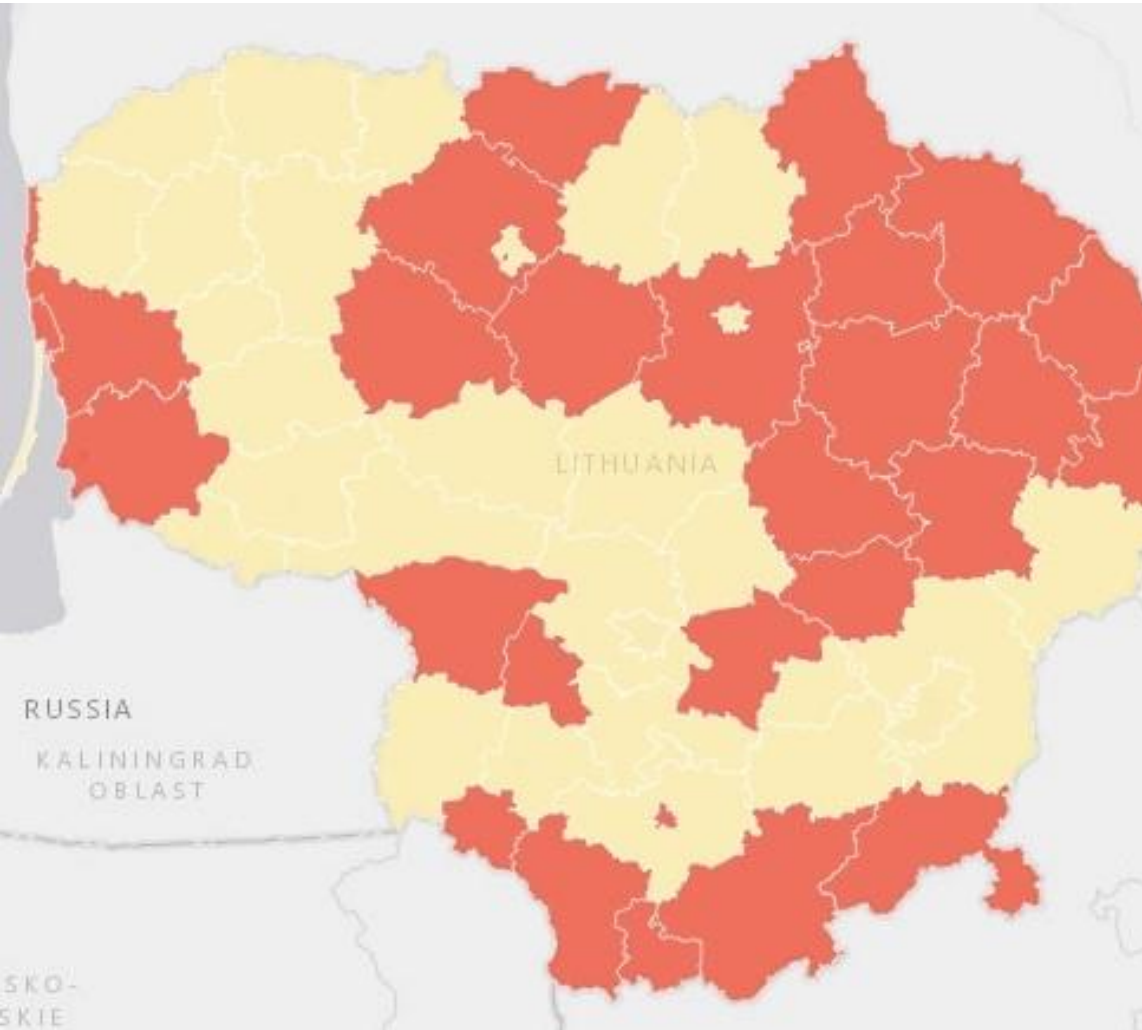






p		TWMA	STWMA	STUA	Moran's I	Geary's C	Gettis-Ord G
0.1	ACC	0.9245	0.9300	0.9330	0.6950	<b>0.8695</b>	<b>0.8560</b>
	BAC	0.7079	0.7259	0.7383	0.8524	<b>0.8524</b>	<b>0.8620</b>
	GAC	0.6531	0.6788	0.6962	0.8522	<b>0.8522</b>	<b>0.8697</b>
0.2	ACC	0.8940	0.8930	0.8965	<b>0.8375</b>	<b>0.8455</b>	<b>0.8280</b>
	BAC	0.8041	0.8045	0.8017	<b>0.8405</b>	<b>0.8454</b>	<b>0.8326</b>
	GAC	0.7906	0.7913	0.7866	<b>0.8405</b>	<b>0.8454</b>	<b>0.8326</b>
0.3	ACC	<b>0.8565</b>	<b>0.8535</b>	<b>0.8545</b>	<b>0.8395</b>	<b>0.8370</b>	<b>0.8370</b>
	BAC	<b>0.8155</b>	<b>0.8151</b>	<b>0.8188</b>	<b>0.8329</b>	<b>0.8277</b>	<b>0.8329</b>
	GAC	<b>0.8081</b>	<b>0.8085</b>	<b>0.8132</b>	<b>0.8327</b>	<b>0.8273</b>	<b>0.8380</b>
0.4	ACC	<b>0.8360</b>	<b>0.8385</b>	<b>0.8400</b>	<b>0.8405</b>	<b>0.8400</b>	<b>0.8370</b>
	BAC	<b>0.8251</b>	<b>0.8264</b>	<b>0.8291</b>	<b>0.8401</b>	<b>0.8387</b>	<b>0.8380</b>
	GAC	<b>0.8232</b>	<b>0.8240</b>	<b>0.8271</b>	<b>0.8401</b>	<b>0.8386</b>	<b>0.8379</b>
0.5	ACC	<b>0.8335</b>	<b>0.8410</b>	<b>0.8370</b>	<b>0.8330</b>	<b>0.8395</b>	<b>0.8380</b>
	BAC	<b>0.8336</b>	<b>0.8411</b>	<b>0.8372</b>	<b>0.8332</b>	<b>0.8387</b>	<b>0.8382</b>
	GAC	<b>0.8335</b>	<b>0.8411</b>	<b>0.8371</b>	<b>0.8331</b>	<b>0.8395</b>	<b>0.8381</b>
0.6	ACC	<b>0.8380</b>	<b>0.8420</b>	<b>0.8425</b>	<b>0.8385</b>	<b>0.8375</b>	<b>0.8380</b>
	BAC	<b>0.8236</b>	<b>0.8271</b>	<b>0.8273</b>	<b>0.8356</b>	<b>0.8346</b>	<b>0.8355</b>
	GAC	<b>0.8209</b>	<b>0.8243</b>	<b>0.8244</b>	<b>0.8355</b>	<b>0.8395</b>	<b>0.8354</b>
0.7	ACC	<b>0.8670</b>	<b>0.8675</b>	<b>0.8690</b>	<b>0.8465</b>	<b>0.8450</b>	<b>0.8355</b>
	BAC	<b>0.8350</b>	<b>0.8358</b>	<b>0.8402</b>	<b>0.8546</b>	<b>0.8545</b>	<b>0.8477</b>
	GAC	<b>0.8312</b>	<b>0.8321</b>	<b>0.8372</b>	<b>0.854</b>	<b>0.8542</b>	<b>0.8472</b>
0.8	ACC	0.8686	0.8750	0.8800	<b>0.8155</b>	<b>0.8180</b>	<b>0.8305</b>
	BAC	0.7535	0.7680	0.7787	<b>0.8327</b>	<b>0.8352</b>	<b>0.8335</b>
	GAC	0.7291	0.7473	0.7605	<b>0.8322</b>	<b>0.8347</b>	<b>0.8335</b>
0.9	ACC	0.9275	0.9290	0.9320	<b>0.8325</b>	<b>0.8405</b>	<b>0.8325</b>
	BAC	0.6930	0.6984	0.7070	<b>0.8313</b>	<b>0.8312</b>	<b>0.8290</b>
	GAC	0.6290	0.6372	0.6495	<b>0.8312</b>	<b>0.8311</b>	<b>0.8290</b>

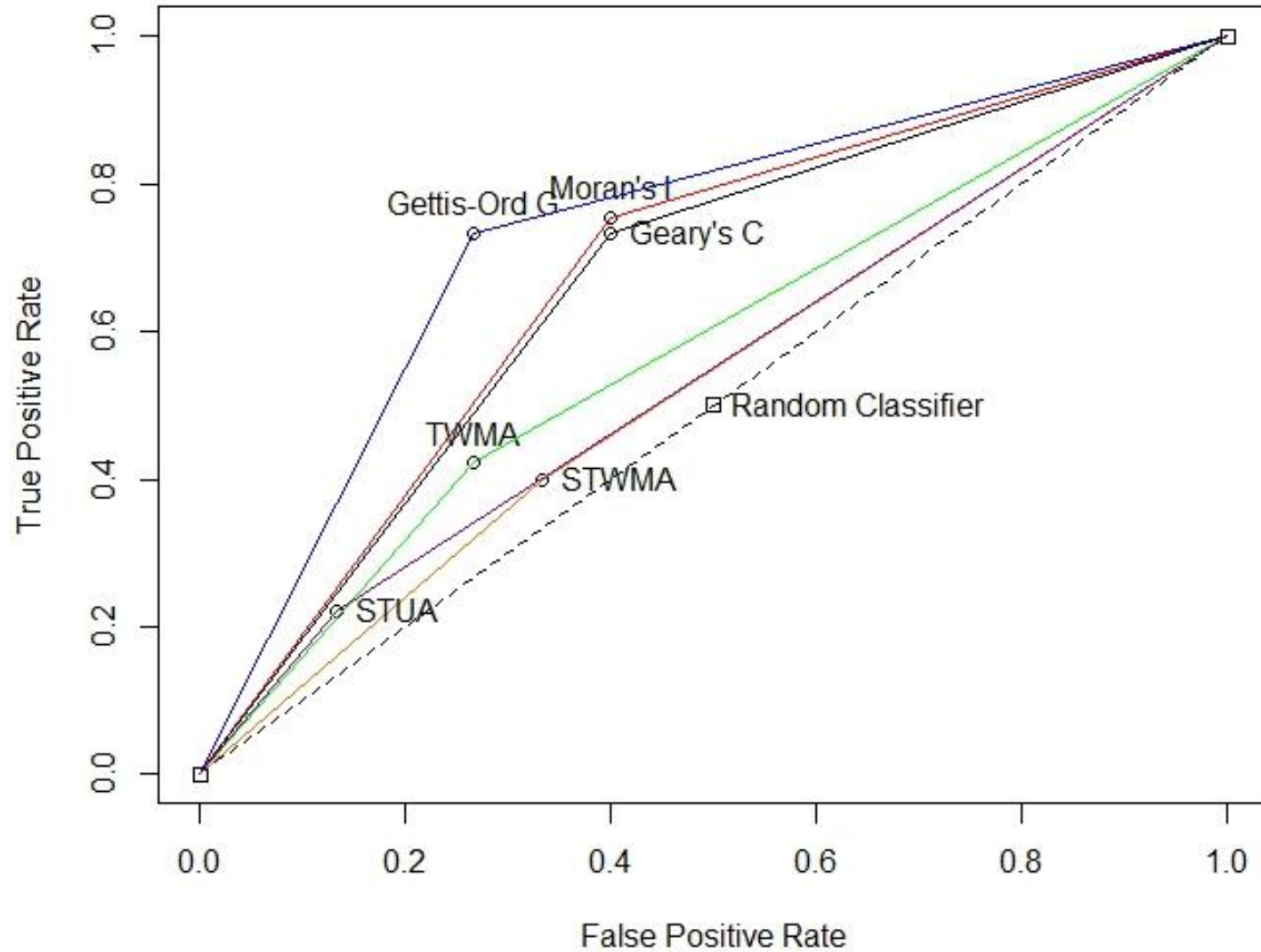
# Gauti rezultatai (2)



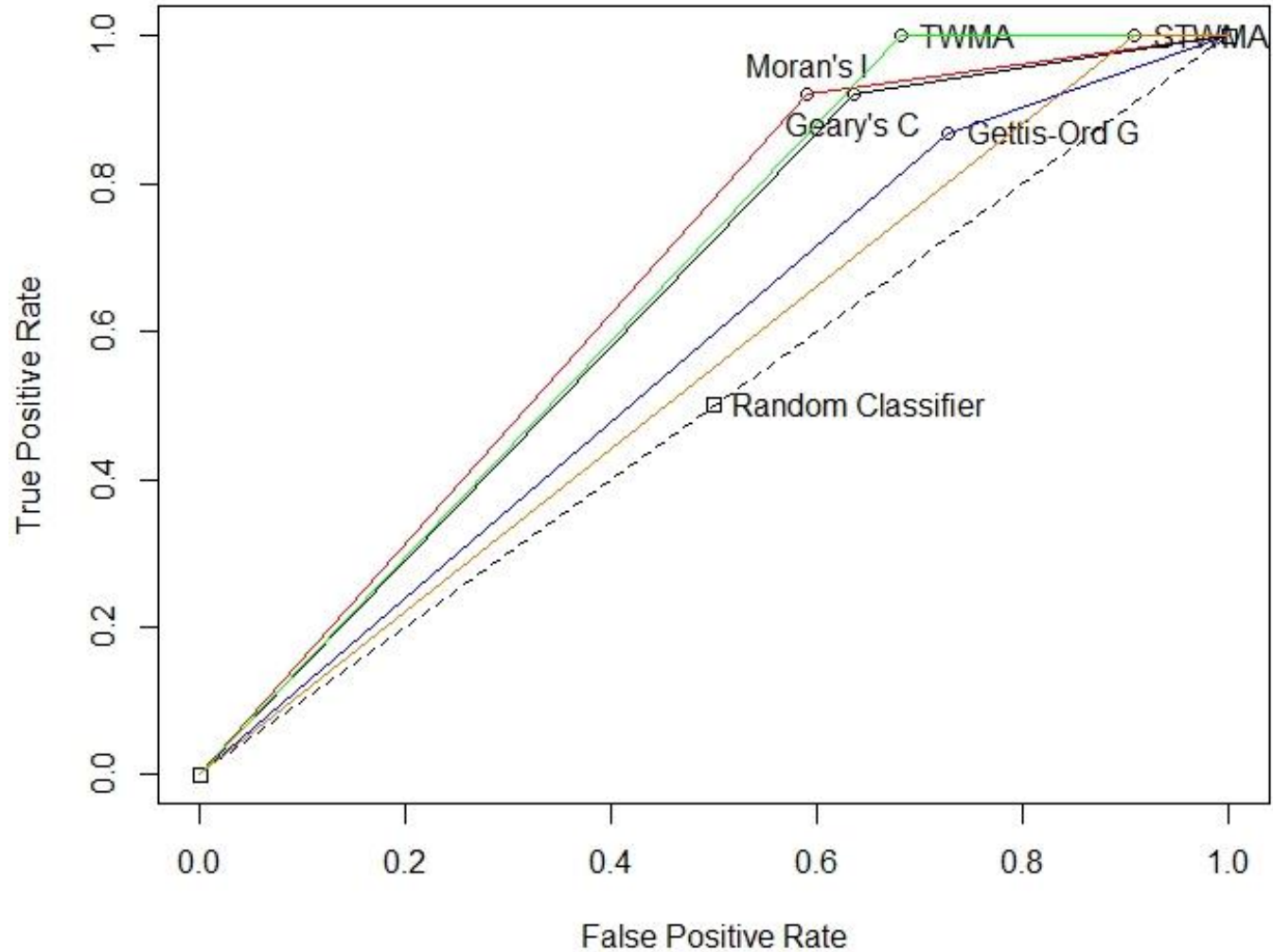
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
ACE	2	2,5	2,2	2,2	1,6	4,5	2,2	4	2,3	2,2	2,3	1,5	1,4	1,1	2	<b>0,9</b>	<b>0,7</b>	<b>0,5</b>	<b>0,3</b>
MN	<b>0,2</b>	<b>0,3</b>	<b>0,2</b>	<b>0,2</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,2</b>	<b>0,2</b>	<b>0,3</b>	<b>0,3</b>	<b>0,1</b>	<b>0,3</b>	<b>0,3</b>	<b>0,3</b>	<b>0,3</b>	<b>0,3</b>	<b>0,6</b>
CSD	7,6	5	3,3	7,6	5	5	3,6	5	5,7	6,5	5,7	19	9	7,6	7,6	5,7	5	5,7	3,3

$$IR_t = \frac{n_1^t}{n_0^t}$$

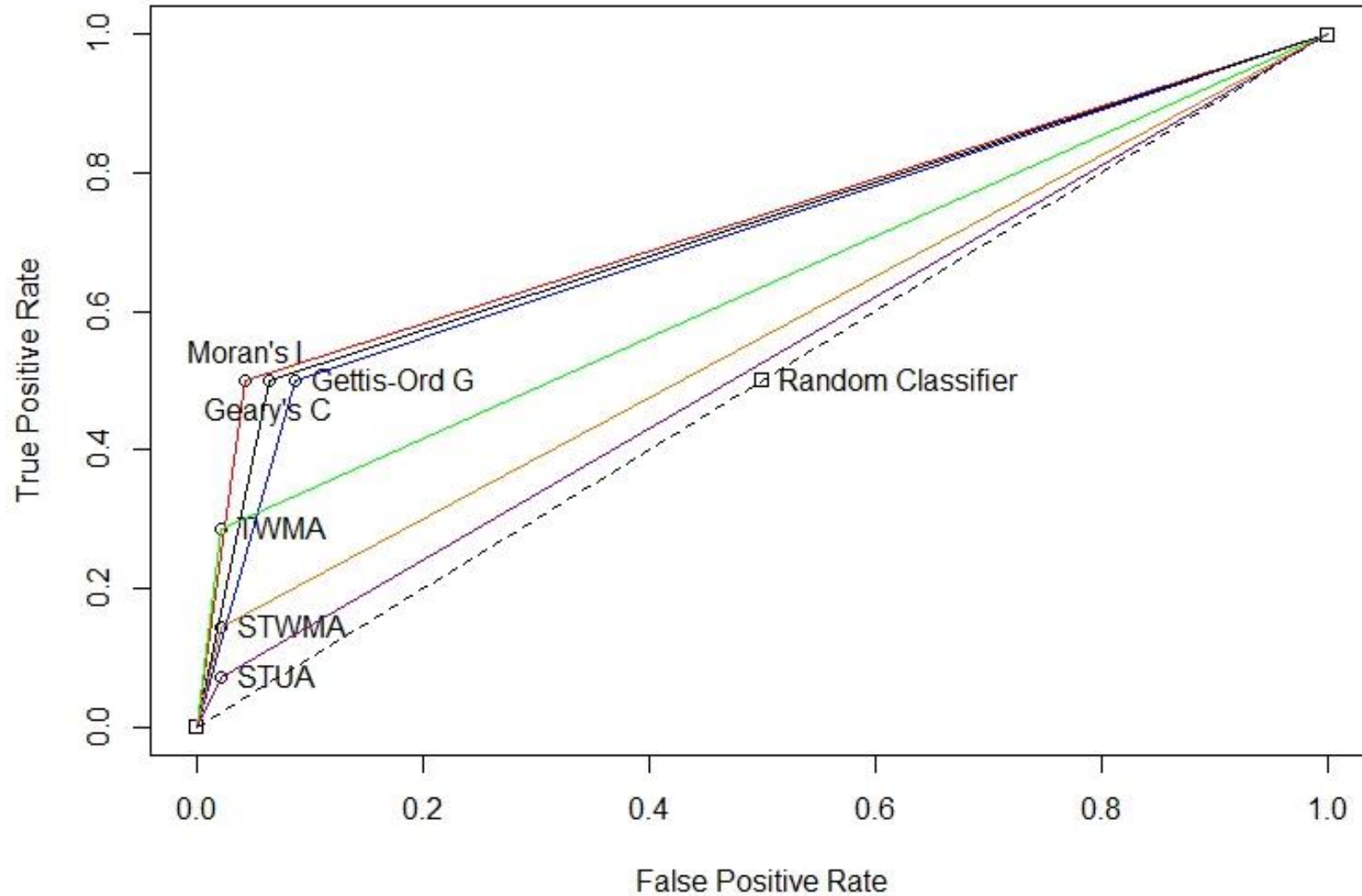
		TWMA	STWMA	STUA	Moran's I	Geary's C	Gettis-Ord G
ACE	ACC	0.5000	0.4667	0.3833	0.7167	0.7000	<b>0.7333</b>
	BAC	0.5778	0.5333	0.5444	0.6778	0.6667	<b>0.7333</b>
	GAC	0.5564	0.5164	0.4389	0.6733	0.6633	<b>0.7333</b>



		TWMA	STWMA	STUA	Moran's I	Geary's C	Gettis-Ord G
MN	ACC	0.7500	0.6667	0.6667	<b>0.7333</b>	0.6500	0.7167
	BAC	0.6591	0.5455	0.5455	<b>0.6651</b>	0.6423	0.5706
	GAC	0.5641	0.3015	0.3015	<b>0.6138</b>	0.5787	0.4867



		TWMA	STWMA	STUA	Moran's I	Geary's C	Gettis-Ord G
CSD	ACC	0.8167	0.7833	0.7667	<b>0.8500</b>	0.8333	0.8167
	BAC	0.6320	0.5606	0.5278	<b>0.7283</b>	0.7174	0.7065
	GAC	0.5287	0.3738	0.2643	<b>0.6956</b>	0.6837	0.6757





# Išvados

1. Generuotiems duomenims, kai klasių žymių binominio skirstinio  $p = 0.3 - 0.7$ , naudojami visi 6 nagrinėti metodai yra tinkami klasifikavimui.
2. Nagrinėjamus duomenis (mirtingumą) geriausiai klasifikuoja metodai:
  - a) ACE (ūminio kardiovaskulinio įvykio) atveju – Gettis-Ord G;
  - b) MN (piktybinių navikų) atveju – Moran's I;
  - c) CSD (kraujotakos sistemos ligų) atveju – Moran's I.

Ačiū už dėmesį