

LATVIJAS UNIVERSITĀTE
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE
MATEMĀTIKAS NODAĻA

**DAUDZDIMENSIJU SADALĪJUMU LIETOŠANAS
ASPEKTI MULTISPEKTRĀLU ATTĒLU ANALĪZĒ**

MAĢISTRA DARBS

Autors: **Mārtiņš Puķītis**

Stud. apl. mp04314

Darba vadītājs: Aivars Lorencs

RĪGA 2015

Anotācija

Darbs ir koncentrēts uz klasifikācijas iespēju analīzi, balstoties uz hipotēzi par daudzdimensionālu sadalījumu, datiem ar dažādu izšķirtspēju. Analīzes piemēram tika izvēlēts uzdevums no 2014. gada IEEE GRSS Data Fusion Contest. Augstas izšķirtspējas RGB attēls un zemākas izšķirtspējas termālie infrasarkanie hiperspektrālie dati, kas iegūti no pilsētas teritorijas, tika apstrādāti, lai iegūtu klasifikāciju katram augstas izšķirtspējas pikselim. Darbā tika analizēta vispārinātā normālā, gamma un Dirihlē sadalījuma lietošana dažādiem klasifikatoriem.

Atslēgvārdi

Tālizpēte, hiperspektrāli attēli, attēlu klasifikācija, datu kombinēšana, daudzdimensionāli sadalījumi.

Abstract

This paper is focused on the analysis of classification possibilities of data with different spatial resolutions based on hypotheses of different types of multivariate distributions. The task set up for the 2014 IEEE GRSS Data Fusion Contest was chosen as an application example. High resolution RGB image and lower resolution thermal infrared image from the same urban area were processed to perform classification of each higher resolution pixel. Application of generalized normal, gamma and Dirichlet distribution were analyzed for different types of classifiers.

Index Terms

Remote sensing, hyperspectral imaging, image classification, data fusion, multidimensional distributions.

SATURS

Ievads.....	4
NODAĻA 1. Daudzdimensionāli sadalījumi.....	5
1.1 Normālais sadalījums	5
1.2 Gamma sadalījums	6
1.3 Dirihlē sadalījums	7
NODAĻA 2. Attēlu klasifikācija	8
2.1 Problēmas apraksts	8
2.2 Klasifikācijas metodes	9
2.3 Vispārinātais normālais sadalījums	10
2.4 Gamma sadalījums	15
2.5 Dirihlē sadalījums	18
2.6 Funkciju kombinācijas	22
Secinājumi	29
Izmantotā literatūra un avoti.....	30

Ievads

Tālizpētes dati, kas iegūti no lidmašīnām un satelītiem, tiek plaši lietoti pilsētvides plānošanā, dabas resursu pārvaldībā, lauksaimniecībā un citās jomās. Dažādu veidu sensori, piemēram, multispektrālie, hiperspektrālie, LiDAR, SAR tiek izmantoti, lai iegūtu dažāda veida datus. Bieži vien vienas teritorijas pētīšanai tiek vienlaikus izmantoti vairāku veidu sensori. Piemēram, lai veiktu meža inventarizāciju, lieto LiDAR kombinācijā ar multispektrāliem datiem no viena un tā paša meža apgabala. LiDAR datus izmanto, lai uzzinātu koku augstumu, savukārt no multispektrālajiem datiem var noteikt sugas vai novērtēt koku veselības stāvokli [1]. Parasti šādos gadījumos no dažādiem avotiem saņemtos datus mēdz kombinēt, lai risinātu konkrētu uzdevumu. Dati, kas saņemti no optiskajiem sensoriem, parasti ir iegūti trīsdimensionāla attēla formā, kur katrs pikselis ir saistīts ar telpiskajām koordinātām, kas tajā pašā laikā ir iegūtas no GPS. Pikseļu izmērs šādā situācijā ir atkarīgs no attāluma līdz pētāmajam objektam, aplūkošanas leņķa un izmantotā sensora jutības un to parasti var izvēlēties datu apstrādes laikā.

Ja iegūtie dati ir ar atšķirīgu izšķirtspēju, to apvienošana ir ievērojami sarežģītāka. Ir nepieciešams saistīt iegūtos attēlus ar ģeogrāfiskajām koordinātām, jo citādi šādu datu kombinēšana kļūst neiespējama.

Viens no galvenajiem tālizpētes uzdevumiem ir klasificēt kāda ģeogrāfiska reģiona attēla pikselus. Ja katru klasi reprezentējošos datus var uztvert kā kāda daudzdimensionāla sadalījuma izlasi, tad to klasifikācijai var lietot Beijesa klasifikatoru. Ar šādu metodi ir iegūti labi rezultāti balstoties uz daudzdimensionālo normālo sadalījumu [2].

Šī darba mērķis ir izpētīt šāda veida klasifikāciju, lietojot citus daudzdimensionālos sadalījumus un salīdzināt iegūtos rezultātus ar atbilstošajiem rezultātiem pie normālā sadalījuma. Klasifikācija tiks pielietota datiem no 2014. gada IEEE GRSS Data Fusion Contest (DFC) [3]. Augstas izšķirtspējas ($\sim 0,2m \times 0,2m$ pikselis) RGB attēls un zemākas izšķirtspējas ($\sim 1m \times 1m$ pikselis) termālie infrasarkanie (TI) hiperspektrālie dati 84 joslās ar viļņa garumiem no $7,8 \mu m$ līdz $11,5 \mu m$, kas iegūti no pilsētas teritorijas.

NODAĻA 1. DAUDZDIMENSIONĀLI SADALĪJUMI

1.1 Normālais sadalījums

Dots: U_1, \dots, U_n neatkarīgi gadījumlīkumi ar standartnormālo sadalījumu. Šo sadalījumu kopējā sadalījuma blīvumfunkcija ir:

$$p_U(u) = (2\pi)^{-\frac{n}{2}} \exp \left[-\frac{1}{2} \sum_{j=1}^n u_j^2 \right] = (2\pi)^{-\frac{n}{2}} \exp \left[-\frac{1}{2} u^T u \right]. \quad 1.1.$$

Izmantojam nesingulāru lineāru transformāciju

$$Y = HU, \quad \text{kur } |H| \neq 0$$

Tad $Y = (Y_1, \dots, Y_n)^T$ kopējā sadalījuma blīvumfunkcija ir:

$$p_Y(y) = (2\pi)^{-\frac{n}{2}} |H^{-1}| \exp \left[-\frac{1}{2} y^T (H^T)^{-1} H^{-1} y \right] = (2\pi)^{-\frac{n}{2}} |A|^{\frac{1}{2}} \exp \left[-\frac{1}{2} y^T A y \right], \quad 1.2.$$

kur $A = (H^T)^{-1} H^{-1}$ ir pozitīvi definīta matrica. Šis ir daudzdimensionālu normālā sadalījuma speciālgadījums ar $E(Y) = 0$. Sadalījuma kovariāciju matrica ir:

$$\begin{aligned} \text{Var}(Y) &= E[HUU^T H^T] \\ &= HE[UU^T]H^T \\ &= HH^T \\ &= A^{-1}. \end{aligned} \quad 1.3.$$

Lai iegūtu vispārīgo gadījumu, aplūkosim $Z = (Z_1, \dots, Z_n)^T$ kopējo sadalījumu, kur $Z = H(U + \mu)$:

$$p_Z(z) = (2\pi)^{-\frac{n}{2}} |A|^{\frac{1}{2}} \exp \left[-\frac{1}{2} (z - \mu)^T A (z - \mu) \right]. \quad 1.4.$$

Šajā gadījumā $E(Z) = \mu$ un $\text{Var}(Z) = \text{Var}(Y) = A^{-1}$ [4].

Definīcija.

$X = (X_1, \dots, X_n)^T$ sauc par sadalītu pēc vispārinātā normālā sadalījuma, ja tā sadalījuma blīvumfunkcija ir:

$$p_X(x) = \frac{n \Gamma \left(\frac{n}{2} \right)}{2^{\frac{n+k}{k}} \pi^{\frac{n}{2}} \Gamma \left(\frac{n+k}{k} \right) \sqrt{|\Sigma|}} * \exp \left[-\{(x - \mu)^T \Sigma^{-1} (x - \mu)\}^{\frac{k}{2}} \right], \quad 1.5.$$

kur $\mu \in \mathbb{R}^n$, Σ ir $n \times n$ pozitīvi definīta matrica un $k > 0$.

$$E(X) = \mu, \quad 1.6.$$

$$\text{cov}(X) = \frac{2^{\frac{2}{k}} \Gamma\left(\frac{n+2}{k}\right)}{n \Gamma\left(\frac{n}{2}\right)} \Sigma. \quad 1.7.$$

Jāatzīmē, ka vispārinātais normālais sadalījums patiesībā ir vesela sadalījumu *klase*. Katru specifisko sadalījumu iegūst, izvēloties parametru k , kas rāda sadalījuma blīvumfunkcijas ‘izliekumu’. Acīmredzams, ka, izvēloties $k = 2$ iegūst daudzdimensionālo normālo sadalījumu [4][5].

1.2 Gamma sadalījums

Modelis.

Detaļa tiek uzstādīta laikā $X_0 = 0$ un, kad tā salūst, to aizstāj ar identisku (vai līdzīgu detaļu). Kad salūst jaunā detaļa, to atkal aizstāj u.t.t.. Apzīmē i -tās detaļas darbības laiku ar Y_i un laiku, kad ir jāaizstāj i -tā detaļa ar X_i . Acīmredzami, $X_i = X_{i-1} + Y_i$.

Konstrukcija.

Aplūko n -dimensionālu gadījumlielumu $Y = (Y_1, \dots, Y_n)^T$. Visi Y_i ir savstarpēji neatkarīgi un katrs Y_i ir gamma sadalīts ar parametriem $\alpha_i > 0, \gamma_i \in \mathbb{R}$ un $\beta > 0$ jeb $Y_i \sim G(\alpha_i, \gamma_i, \beta)$. Veic gadījumlieluma X konstrukciju, balstoties uz iepriekš aplūkoto modeli:

$$X_i = X_{i-1} + Y_i = \sum_{j=1}^i Y_j, \quad \forall i = 1, \dots, n. \quad 1.8.$$

Iegūtais $X = (X_1, \dots, X_n)^T$ ir ar daudzdimensionālu gamma sadalījumu un tā sadalījuma blīvumfunkcija ir:

$$p_X(x) = \frac{(x_1 - \gamma_1)^{\alpha_1 - 1}}{\beta^{\alpha_n^*} \prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=2}^n (x_i - x_{i-1} - \gamma_i)^{\alpha_i - 1} * \exp\left[-\frac{x_n - \sum_{i=1}^n \gamma_i}{\beta}\right], \quad 1.9.$$

kur $\alpha_n^* = \sum_{j=1}^n \alpha_j$.

X_i robežsadalījums ir $G(\alpha_i^*, \gamma_i^*, \beta)$, kur $\alpha_i^* = \sum_{j=1}^i \alpha_j$ un $\gamma_i^* = \sum_{j=1}^i \gamma_j$.

$$E[X_i] = \beta \alpha_i^* + \gamma_i^* \quad 1.10.$$

$$\text{var}(X_i) = \beta^2 \alpha_i^* \quad 1.11.$$

$$\text{cov}(X_i, X_j) = \text{var}(X_i) = \beta^2 \alpha_i^*, \quad \text{ja } i < j \quad 1.12.$$

$$\rho = \text{corr}(X_i, X_j) = \sqrt{\frac{\alpha_i^*}{\alpha_j^*}}, \quad \text{ja } i < j \text{ [4]} \quad 1.13.$$

1.3 Dirihlē sadalījums

X_0, X_1, \dots, X_m ir neatkarīgi gadījumielumi un X_i ir ar χ^2 sadalījumu ar v_i brīvības pakāpēm ($v_i > 0$, bet ne obligāti vesels skaitlis). Konstruē

$$Y_j = \frac{X_j}{\sum_{i=0}^m X_i}, \quad j = 1, 2, \dots, m. \quad 1.14.$$

Kopējā blīvuma f-ja ir

$$p_{Y_1, \dots, Y_m}(y_1, \dots, y_m) = \frac{\Gamma(\sum_{j=0}^m \theta_j)}{\prod_{j=0}^m \Gamma(\theta_j)} \left[\prod_{j=1}^m y_j^{\theta_j-1} \right] \left(1 - \sum_{j=1}^m y_j \right)^{\theta_0-1} \quad 1.15.$$

kur $\theta_j = \frac{v_j}{2}$.

$$E(Y_i) = \frac{\theta_i}{\Theta}, \quad i = 1, 2, \dots, m, \quad 1.16.$$

$$\text{kur } \Theta = \sum_{i=1}^m \theta_i,$$

$$\text{var}(Y_i) = \frac{\theta_i(\Theta - \theta_i)}{\Theta^2(\Theta + 1)} \text{ [4]}. \quad 1.17.$$

NODAĻA 2. ATTĒLU KLASIFIKĀCIJA

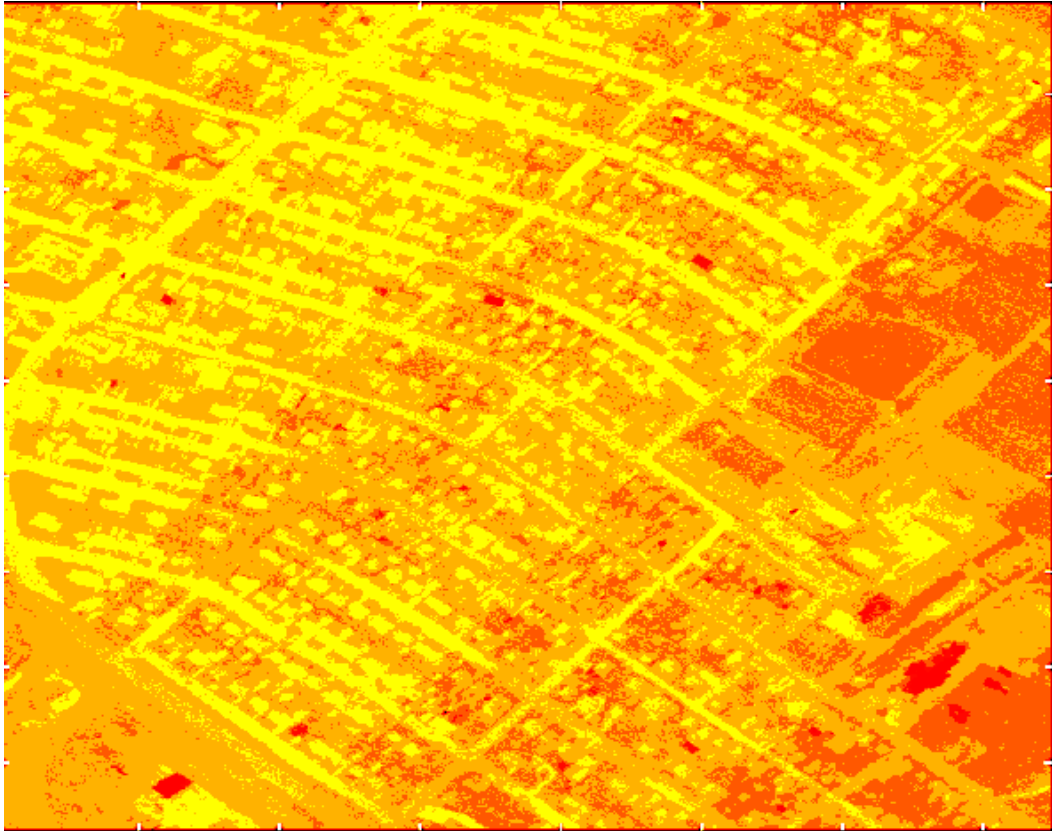
2.1 Problēmas apraksts

Pētījumā tiek izmantoti dati no „IEEE GRSS Data Fusion Contest”. Dots augstas kvalitātes redzamās gaismas attēls un zemākas kvalitātes termālais attēls. Attēli ir uzņemti no lidaparāta, pārlidojot Thetford Mines Kvebekā, Kanādā.

Redzamās gaismas attēls sastāv no trim joslām, kas norāda attiecīgi sarkanās, zaļās un zilās gaismas intensitāti (turpmāk tekstā vizuālais attēls). Šis attēls ir iegūts vairākos pārlidojumos un ir nepilnīgs – starp pārlidojumu joslām ir nenofotografēti laukumi. Termālais attēls sastāv no 84 spektra joslām. Termālais attēls pārklāj visu aplūkoto reģionu.



2.1. att. Pētījuma analizē izmantotais vizuālais attēls



2.2. att. **Pētījuma analizē izmantotais termālais attēls (3 joslu vizuāla reprezentācija)**

No attēla ir izdalītas tā saucamās dizaina kopas, kuras pieder konkrētai klasei. Izdalītās klases ir: ceļi, koki, sarkani jumti, pelēki jumti, betona jumti, zālāji un atklāta augsne. Klasifikācijas uzdevums ir, balstoties uz informāciju no dizaina kopām, sadalīt pārējo attēlu norādītajās klasēs [3].

2.2 Klasifikācijas metodes

Vispārīgs apraksts.

Klasifikāciju veic balstoties uz hipotēzi par pikseļu sadalījumu. Izmantojot no dizaina kopām iegūtos datus, novērtē sadalījuma parametrus katrai klasei no vizuālā un termālā attēla. Aprēķina, kāda ir katra klasificējamā pikseļa sadalījuma blīvumfunkcijas vērtība, ja tas pieder kādai no klasēm. Salīdzina katra pikseļa sadalījuma blīvumfunkcijas vērtības pa visām klasēm. Izvēlas klasi, kurai sadalījuma blīvumfunkcija ir ar maksimālo vērtību. Šo metodi sauc par Beijesa klasifikatoru.

Apzīmējumi:

- $f_k(x)$ – blīvumfunkcija vizuālā attēla pikselim x klasei k ,

- $g_k(y_x)$ – blīvumfunkcija termālā attēla pikselim y_x klasei k ,
- y_x – termālā attēla pikselis, kas atbilst pikseli x saturošam vizuālā attēla apgabalam.

Konkrēti klasifikācijas varianti.

I. Izvēlas klasi, kurai blīvumfunkciju summa ir maksimālā

$$f_k(x) + g_k(y_x) \geq f_j(x) + g_j(y_x), \quad \forall j = 1, \dots, 7. \quad 2.1.$$

II. Piekārto „svarus” iepriekš aplūkotajai formulai

$$\frac{2}{3}f_k(x) + \frac{1}{3}g_k(y_x) \geq \frac{2}{3}f_j(x) + \frac{1}{3}g_j(y_x), \quad \forall j = 1, \dots, 7. \quad 2.2.$$

III. Atrod atbilstošo klasi vizuālajam un termālajam attēlam atsevišķi. Ja tās atšķiras – izvēlas klasi, kurai ir augstāka atbilstošā blīvuma funkcijas vērtība.

$$\begin{aligned} f_k(x) &\geq f_j(x), & \forall j = 1, \dots, 7, \\ g_{k'}(y_x) &\geq f_j(x), & \forall j = 1, \dots, 7, \\ f_k(x) &\geq g_{k'}(y_x) \Rightarrow k, \\ f_k(x) &\leq g_{k'}(y_x) \Rightarrow k'. \end{aligned} \quad 2.3.$$

IV. Dažādu blīvumfunkciju kombināciju lietošana iepriekš aplūkotajās metodēs.

Par attēlu kopumā nekādas drošas informācijas nav, tādēļ tā klasifikāciju var novērtēt tikai vizuāli – iekrāsot vienai klasei piederošos attēla pikseļus vienā krāsā un salīdzināt iegūto „karti” ar oriģinālo attēlu. Skaitliskus klasifikatoru precizitātes novērtējumus var iegūt tikai pielietojot tos dizaina kopām.

Jāatzīmē, ka visu 84 termālo joslu izmantošana klasifikatorā ir ne tikai laikietilpīga, bet arī neefektīva, jo blakusesošās joslas viena no otras būtiski neatšķiras. Pētījumā ir izmantotas septiņas no termālajām joslām.

Joslu atlasē kritērijs: septiņām patvaļīgi izvēlētām joslām aprēķina dizaina kopu vidējās vērtības katrā no klasēm, iegūstot 7×7 matricu; izvēlas to joslu kombināciju, kurai determinants no iegūtās matricas ir lielākais pēc absolūtās vērtības.

2.3 Vispārinātais normālais sadalījums

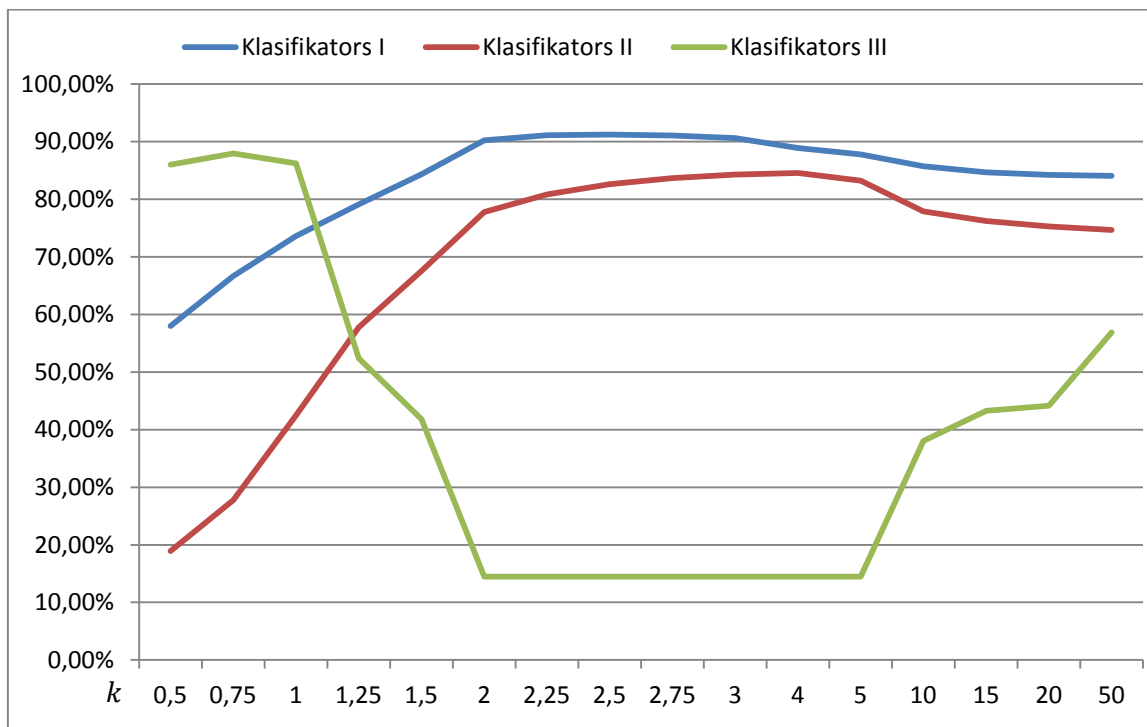
Parametru novērtēšana šim sadalījumam ir triviāla. Sadalījuma blīvumfunkcija ir atkarīga no parametriem μ un Σ :

$$\hat{\mu} = E(X), \quad 2.4.$$

$$\hat{\Sigma} = \frac{n\Gamma\left(\frac{n}{2}\right)}{2^{\frac{n}{2}}\Gamma\left(\frac{n+2}{2}\right)} \text{cov}(X).$$

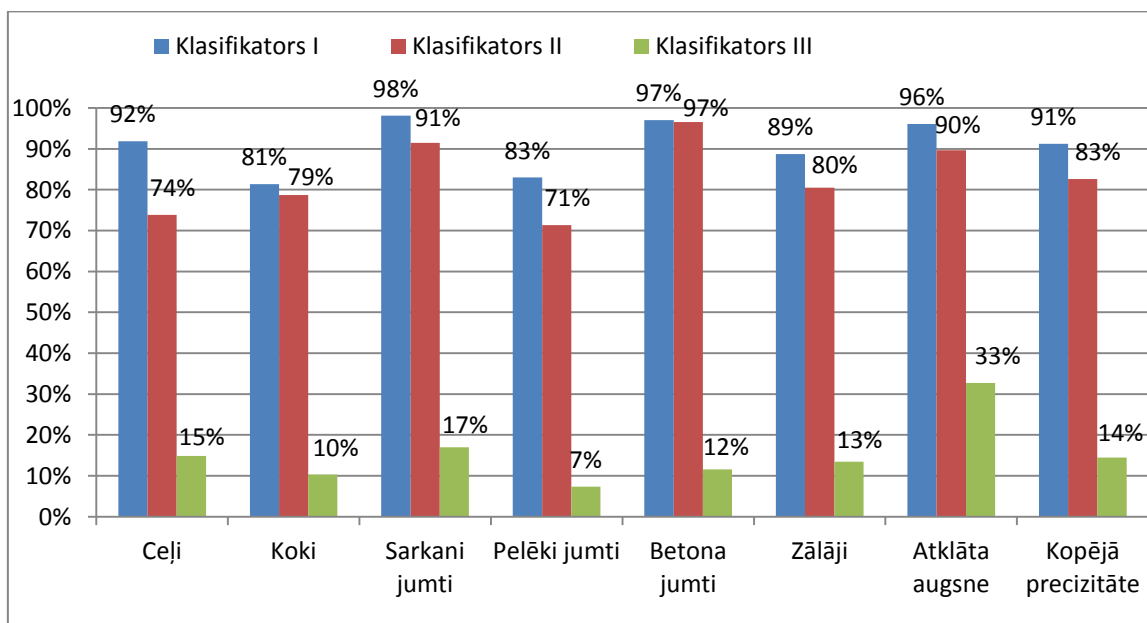
2.5.

Parametrs k ir jāizvēlas. Pētījumā ir salīdzināti klasifikācijas rezultāti dažādām parametra k vērtībām.

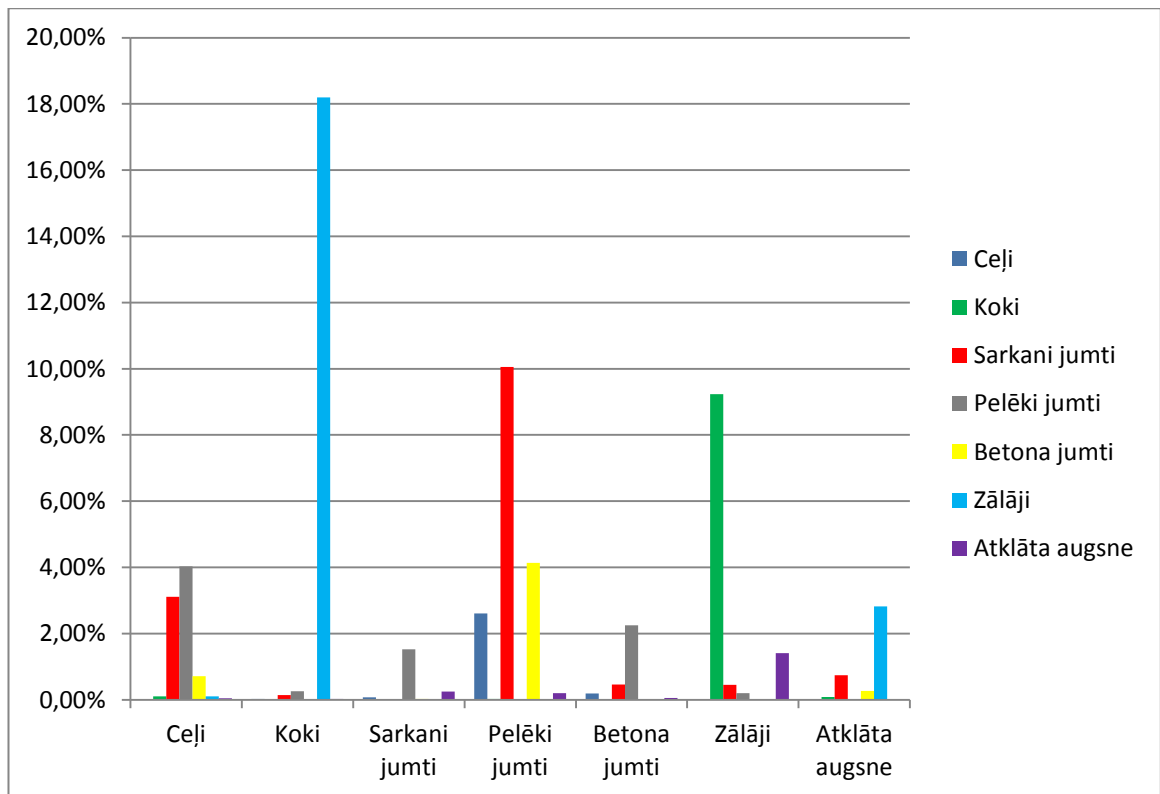


2.3. att. Klasifikatoru kopējā precizitāte vispārinātajam normālajam sadalījumam atkarībā no parametra k vērtības

Visaugstāko kopējo precizitāti dod klasifikators I ar $k = 2,5$. Šai parametra vērtībai tiks aplūkoti detalizētāki klasifikācijas rezultāti.



2.4. att. Klasifikācijas rezultāti vispārinātajam normālajam sadalījumam ar $k=2,5$



2.5. att. Klasifikācijas kļūdas vispārinātajam normālajam sadalījumam ar $k=2,5$, klasifikators I

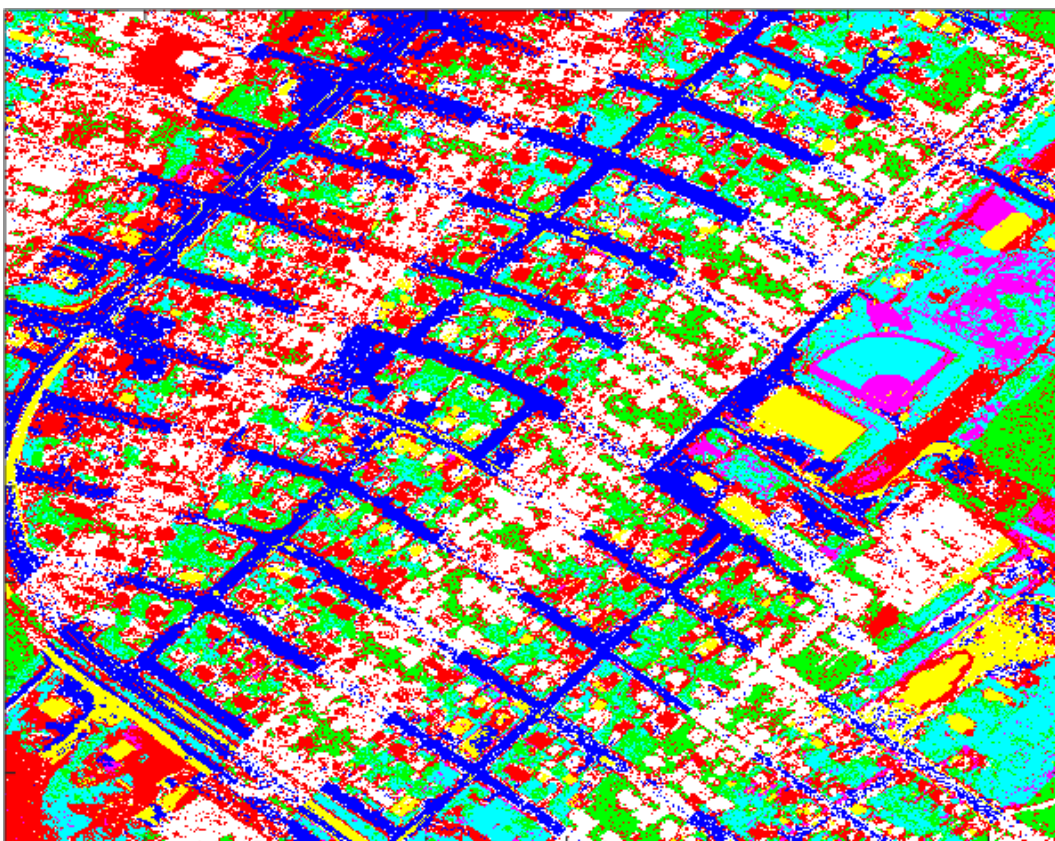
Iegūtās kļūdas īpaši nepārsteidz. Koku un zālāju pikseļi arī vizuāli ir pietiekami līdzīgi, tāpat arī ceļi un pelēkie jumti.

Kā jau nodaļas sākumā tika minēts, klasificēšanu var novērtēt arī vizuāli. Katram aplūkotajam sadalījuma veidam tādēļ tiks arī apskatīta klasifikācijas rezultātā iegūtā karte.

2.1 tabula

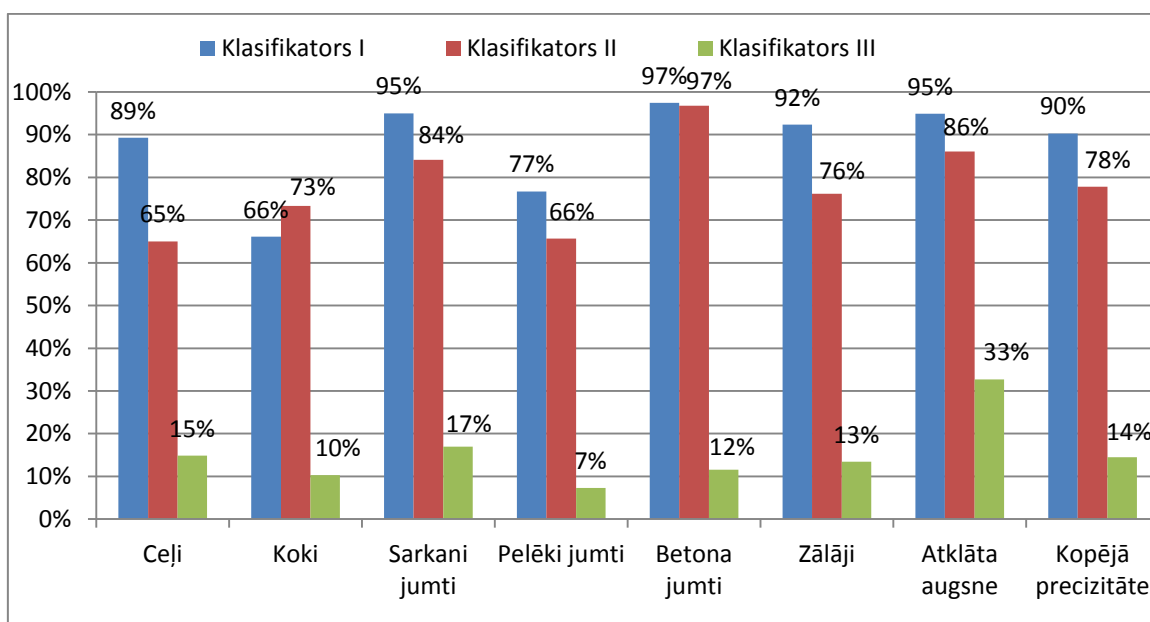
Krāsu atšifrējums klasifikācijas kartēm

	Ceļi
	Koki
	Sarkani jumti
	Pelēki jumti
	Betona jumti
	Zālāji
	Atklāta augsne

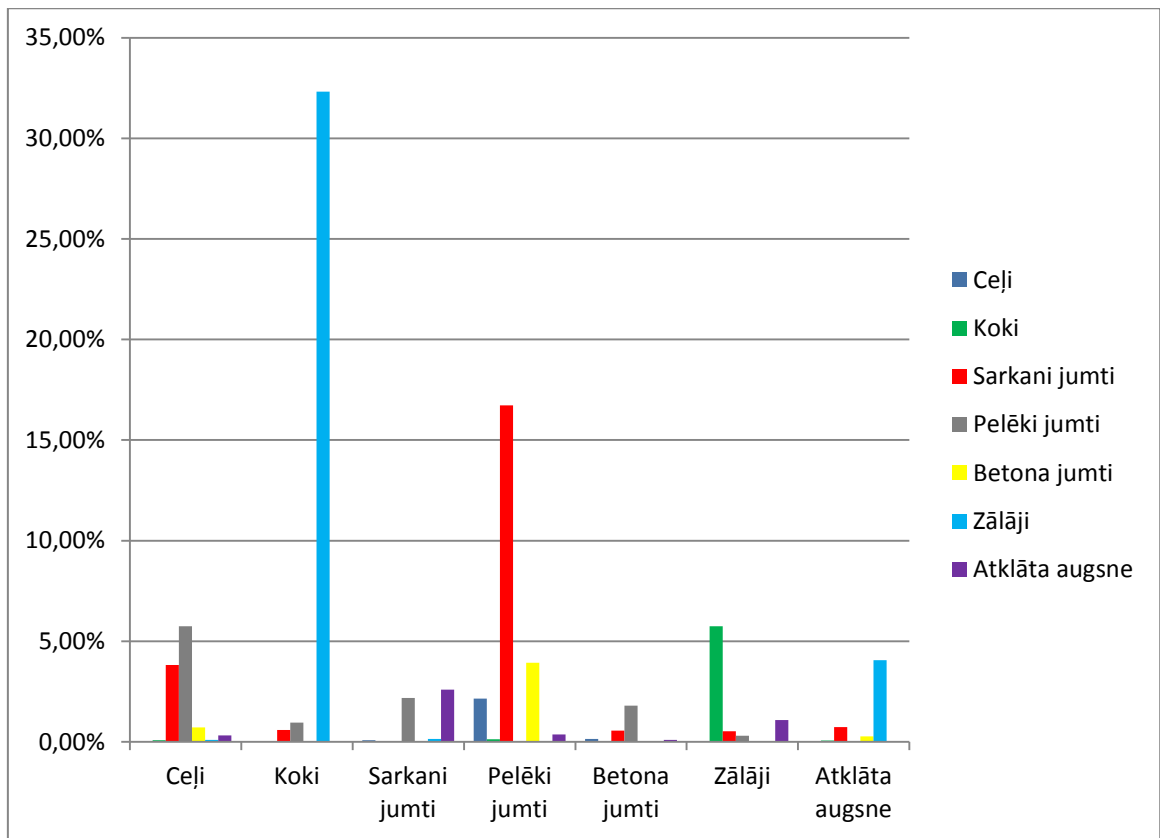


2.6. att. Klasifikācijas karte vispārinātajam normālajam sadalījumam ar $k=2,5$, klasifikators I

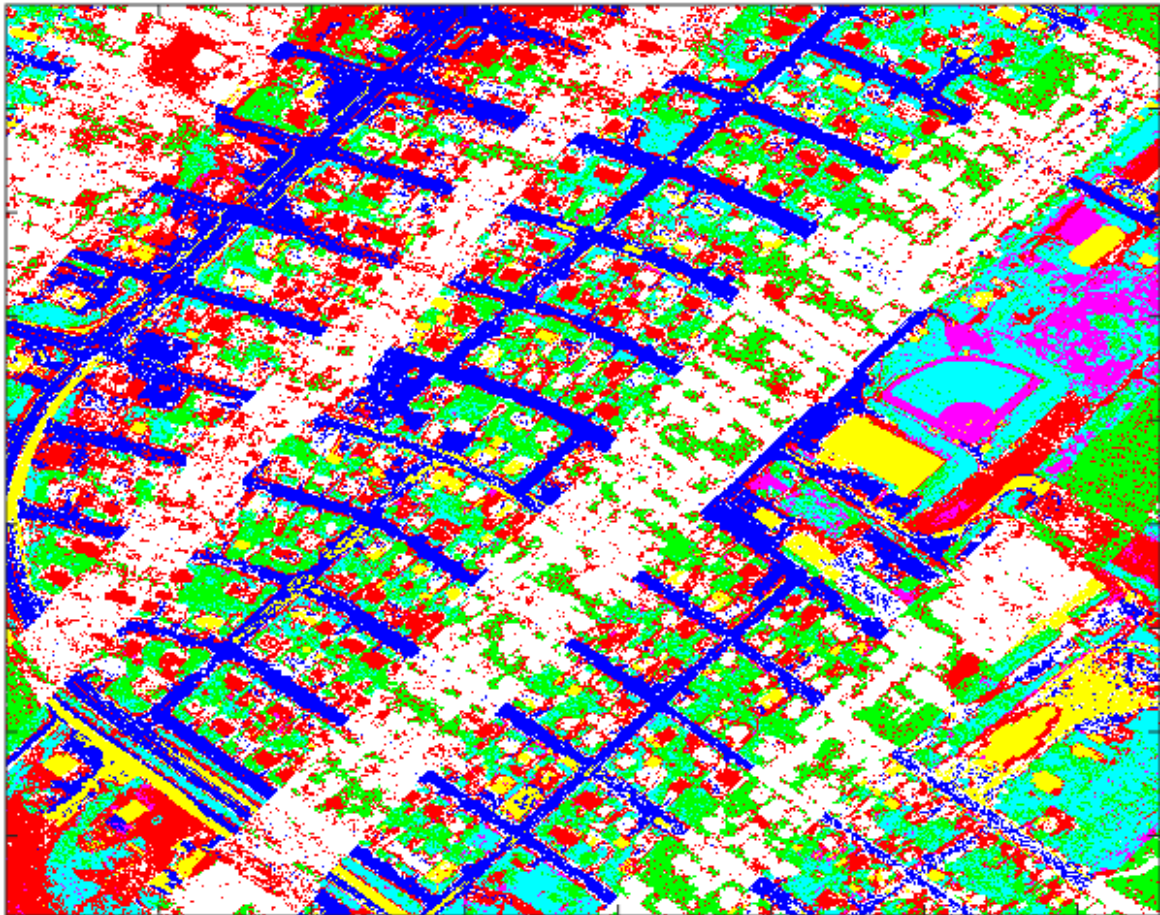
Salīdzināšanai tiks aplūkoti klasifikācijas rezultāti normālajam sadalījumam ($k=2$).



2.7. att. Klasifikācijas rezultāti normālajam sadalījumam



2.8. att. **Klasifikācijas kļūdas normālajam sadalījumam, klasifikators I**



2.9. att. **Klasifikācijas karte normālajam sadalījumam, klasifikators I**

2.4 Gamma sadalījums

Saskaņā ar sadalījuma modeli, katra no vizuālā attēla krāsu joslām (RGB) ir ar gamma sadalījumu, kas atkarīgs no trim parametriem, kur parametrs β visām joslām ir kopīgs. Veic mainīgo transformāciju: $X_1 = R, X_2 = R + G, X_3 = R + G + B$. Analogi notiek mainīgo transformācija termālajam attēlam: apzīmē atlasītās joslas ar $T = (T_1, \dots, T_7)$ un aprēķina:

$$Y_i = \sum_{j=1}^i T_j, \quad i = 1, \dots, 7. \quad 2.6.$$

Rezultātā, sadalījums vizuālā attēla klasēm ir atkarīgs no septiņiem parametriem: $\alpha_1, \alpha_2, \alpha_3, \beta, \gamma_1, \gamma_2, \gamma_3$ (termālā attēla sadalījums atbilstoši ir atkarīgs no piecpadsmit parametriem). Šie parametri ir jāizsaka no konstruēto mainīgo vidējo vērtību un dispersiju novērtējumiem. Pilnībā tos izteikt tomēr nav iespējams – ir $2n + 1$ nezināmie un tikai $2n$ vienādojumi, tādēļ viena no γ_i vērtībām būs jāizvēlas.

Vienkāršības dēļ (tāpat kā viendimensiju gamma sadalījumam), pieņem, ka visi $\gamma_i = 0$. Rezultātā paliek $n + 1$ nezināmie, kas jāizsaka no $2n$ vienādojumiem. Parametru $\hat{\beta}$ var izteikt no jebkura mainīgā vidējās vērtības un dispersijas, par tā novērtējumu $\hat{\beta}$ ņemsim vidējo no atsevišķajiem novērtējumiem.

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n \frac{\text{var}(Z_i)}{E(Z_i)}, \quad 2.7.$$

$$\hat{\alpha}_i^* = \frac{E(Z_i)}{\hat{\beta}}, \quad i = 1, \dots, n, \quad 2.8.$$

$$\hat{\alpha}_i = \hat{\alpha}_i^* - \sum_{j=1}^{i-1} \hat{\alpha}_j, \quad i = 1, \dots, n, \quad 2.9.$$

2.2 tabula

Gamma sadalījuma parametri vizuālajam attēlam

	Ceļi	Koki	Sarkani jumti	Pelēki jumti	Betona jumti	Zālāji	Atklāta augsne
$\hat{\alpha}_1$	5,96	2,46	1,95	0,93	201,99	5,33	9,06
$\hat{\alpha}_2$	5,81	2,87	1,49	0,96	202,17	5,56	7,68
$\hat{\alpha}_3$	6,45	1,94	1,53	1,09	202,50	4,53	7,18
$\hat{\beta}$	0,035	0,042	0,101	0,175	0,002	0,024	0,029

Gamma sadalījuma parametri termālajam attēlam

	Ceļi	Koki	Sarkani jumti	Pelēki jumti	Betona jumti	Zālāji	Atklāta augsne
$\hat{\alpha}_1$	5559	5766	914	3053	821	4641	5872
$\hat{\alpha}_2$	4945	5162	822	2745	732	4154	5242
$\hat{\alpha}_3$	4339	4588	726	2426	649	3682	4632
$\hat{\alpha}_4$	3803	3994	631	2114	568	3214	4012
$\hat{\alpha}_5$	3401	3496	562	1878	498	2815	3539
$\hat{\alpha}_6$	2873	2992	475	1595	429	2423	3058
$\hat{\alpha}_7$	2420	2470	399	1328	356	1996	2543
$\hat{\beta}$	1,01E-08	9,05E-09	6,10E-08	1,81E-08	6,39E-08	1,12E-08	8,88E-09

Sadalījuma blīvumfunkcija ir:

$$p_Z(z) = \frac{z_1^{\hat{\alpha}_1 - 1}}{\hat{\beta}^{\hat{\alpha}_n} \prod_{i=1}^n \Gamma(\hat{\alpha}_i)} \prod_{i=2}^n (x_i - x_{i-1})^{\hat{\alpha}_i - 1} * \exp\left[-\frac{z_n}{\hat{\beta}}\right], \quad 2.10.$$

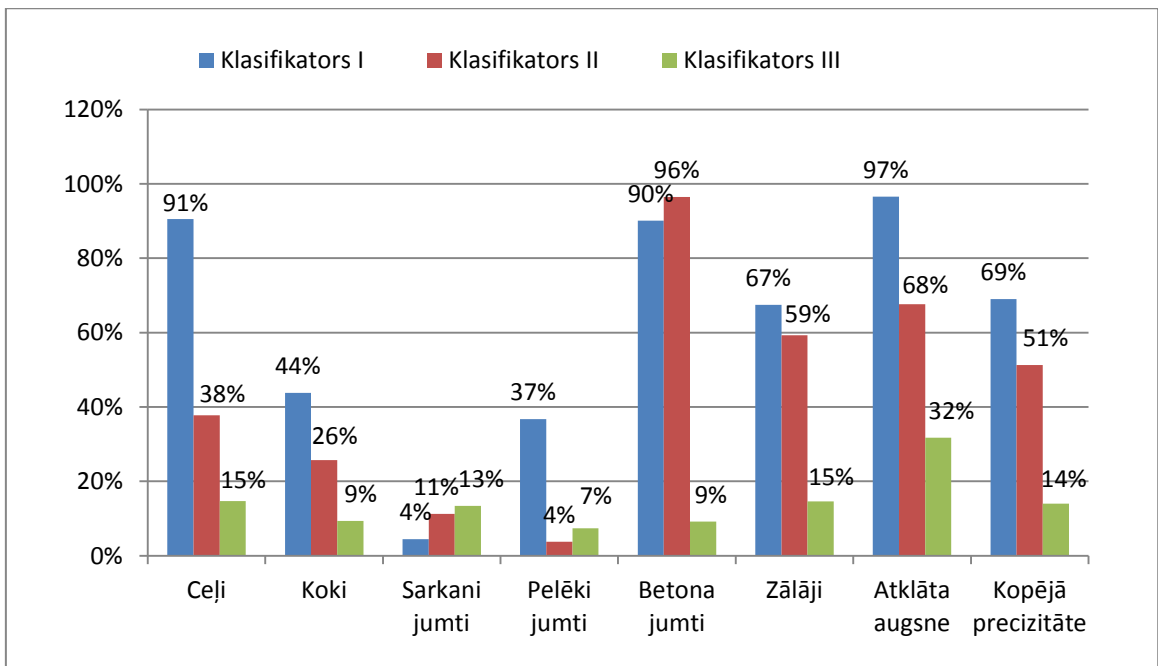
kur Z vietā liek X vai Y atkarībā no aplūkotā attēla.

Šeit jāatzīmē, ka lielām parametru α_i vērtībām rodas tehniska rakstura problēmas blīvumfunkcijas aprēķināšanai. Programma MatLab visiem skaitļiem, kas pārsniedz $1.79e+308$ atgriež „Inf” un turpmāka rēķināšana kļūst neiespējama. Piemēram, $\Gamma(172)$ jau pārsniedz šo limitu. Šo problēmu risina transformējot gamma funkciju jaunā formā un logaritmizējot varbūtību blīvumfunkciju:

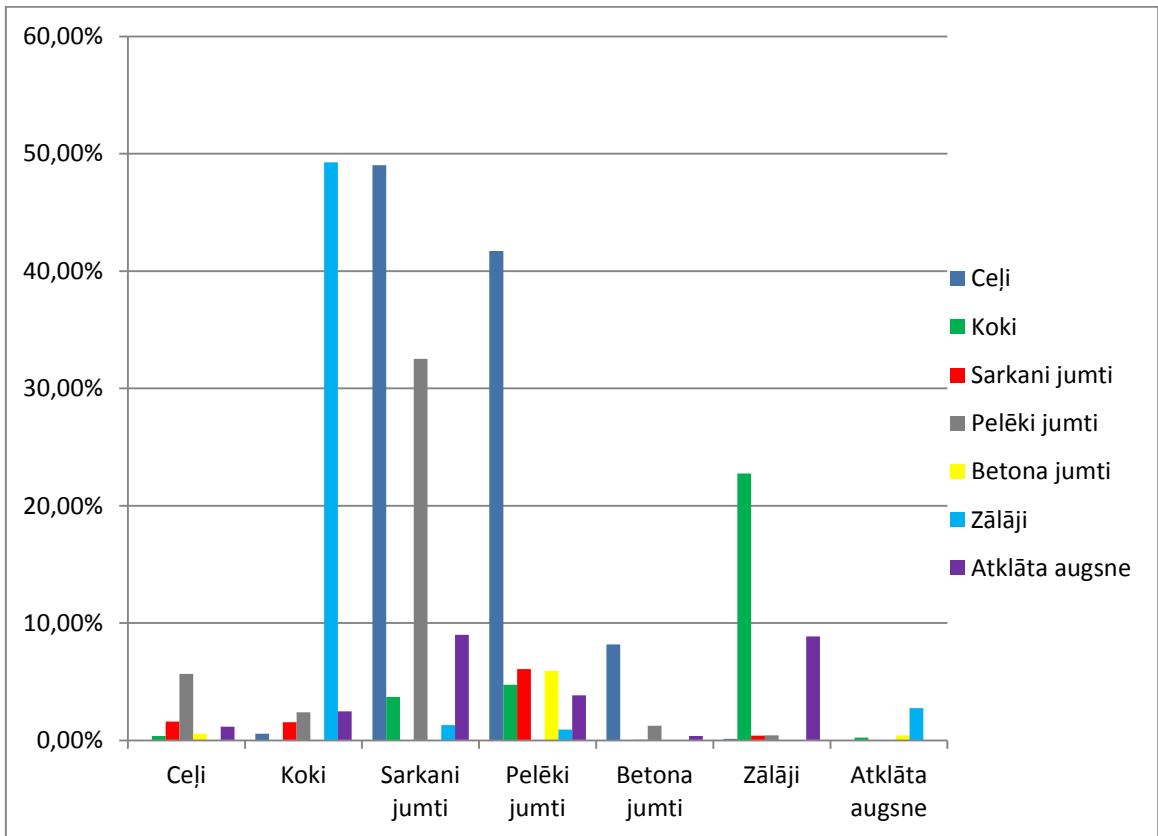
$$\Gamma(t) = \prod_{i=1}^{[t]} (t - i) \Gamma(t - [t]) = t^{[t]} \prod_{i=1}^{[t]} \left(1 - \frac{i}{t}\right) \Gamma(t - [t]) \quad 2.11.$$

$$\begin{aligned} \ln(p_Z(z)) &= \\ &= (\alpha_1 - 1) \ln z_1 + \sum_{i=2}^n (\alpha_i - 1) \ln(z_i - z_{i-1}) - \alpha_n^* \ln \beta - \frac{z_n}{\beta} - \sum_{i=1}^n \ln \Gamma(\alpha_i) = \\ &= (\alpha_1 - 1) \ln z_1 + \sum_{i=2}^n (\alpha_i - 1) \ln(z_i - z_{i-1}) - \alpha_n^* \ln \beta - \frac{z_n}{\beta} - \\ &\quad - \sum_{i=1}^n \left\{ [\alpha_i] \ln(\alpha_i) \sum_{j=1}^{[\alpha_i]} \ln\left(1 - \frac{j}{\alpha_i}\right) + \ln \Gamma(\alpha_i - [\alpha_i]) \right\}. \end{aligned} \quad 2.12.$$

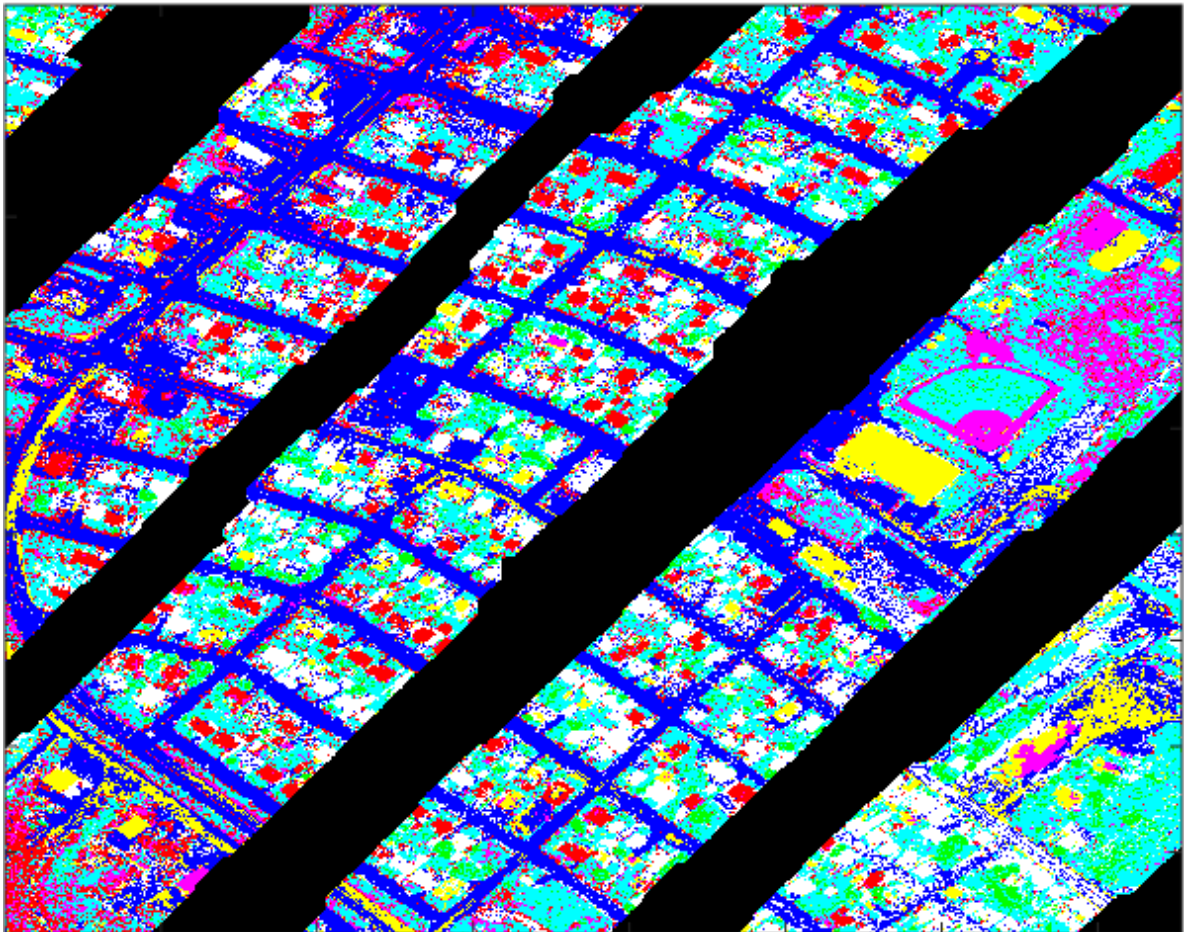
Tā kā klasifikācija balstās uz augstāko blīvumfunkcijas vērtību, funkcijas logaritmizēšana klasifikāciju neietekmē.



2.10. att. **Klasifikācijas rezultāti gamma sadalījumam**



2.11. att. **Klasifikācijas kļūdas gamma sadalījumam, klasifikators I**



2.12. att. **Klasifikācijas karte Gamma sadalījumam, klasifikators I**

2.5 Dirihlē sadalījums

Modelis paredz, ka katra no krāsu joslām ir ar χ^2 sadalījumu. Veic mainīgo transformāciju: $X_1 = \frac{R}{R+G+B}$, $X_2 = \frac{G}{R+G+B}$. Analogi aprēķina transformētos mainīgos termālajam attēlam:

$$Y_i = \frac{T_i}{\sum_{j=1}^7 T_j}, \quad i = 1, \dots, 6. \quad 2.13.$$

Dirihlē sadalījums ir atkarīgs no $n + 1$ parametriem, kur n ir transformēto mainīgo skaits. Vienkāršības labad, apzīmē:

$$\theta = \sum_{i=0}^n \theta_i. \quad 2.14.$$

Parametrus aprēķina no transformēto mainīgo vidējo vērtību un dispersiju novērtējumiem. Šajā gadījumā iegūst $2n$ vienādojumus ar tikai $n + 1$ nezināmajiem, tādēļ parametru vērtības viennozīmīgi nav nosakāmas. θ var izteikt no jebkura atsevišķa mainīgā

vidējās vērtības un dispersijas, kas vispārīgā gadījumā, protams, dos atšķirīgus rezultātus. Par Θ novērtējumu tiks uzskatīta vidējā vērtība no atsevišķajiem novērtējumiem.

$$\hat{\Theta} = \frac{1}{n} \sum_{i=1}^n \frac{E(Z_i)(1 - E(Z_i))}{\text{var}(Z_i)} - 1 \quad 2.15.$$

$$\hat{\theta}_0 = \hat{\Theta} \left(1 - \sum_{i=1}^n E(Z_i) \right) \quad 2.16.$$

$$\hat{\theta}_i = \hat{\Theta} E(Z_i), \quad \text{kur } i = 1, \dots, n.$$

2.4 tabula

Dirihlē sadalījuma parametri vizuālajam attēlam

	Ceļi	Koki	Sarkani jumti	Pelēki jumti	Betona jumti	Zālāji	Atklāta augsne
$\hat{\theta}_0$	3239	179	190	488	17281	283	1079
$\hat{\theta}_1$	2992	223	239	406	17232	331	1366
$\hat{\theta}_2$	2920	261	182	424	17247	349	1159
$\hat{\Theta}$	9150	663	611	1318	51760	963	3605

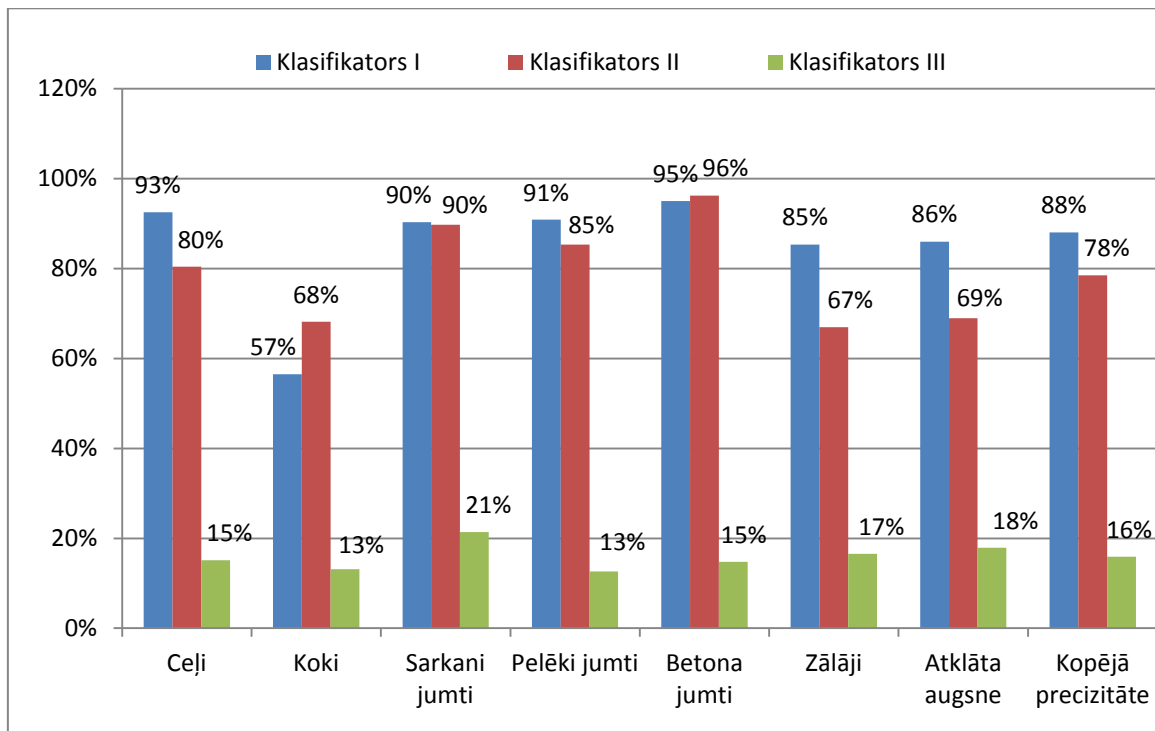
2.5 tabula

Dirihlē sadalījuma parametri termālajam attēlam

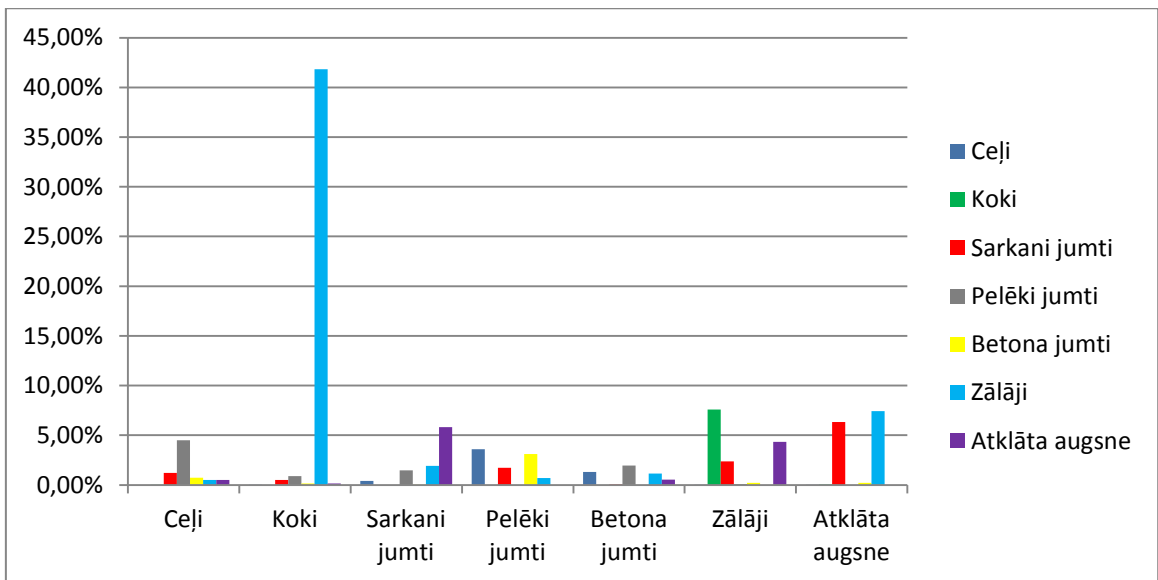
	Ceļi	Koki	Sarkani jumti	Pelēki jumti	Betona jumti	Zālāji	Atklāta augsne
$\hat{\theta}_0$	13003	21375	16817	14806	15845	15408	11296
$\hat{\theta}_1$	29872	49910	38506	34033	36493	35836	26081
$\hat{\theta}_2$	26571	44682	34651	30602	32562	32074	23280
$\hat{\theta}_3$	23316	39713	30575	27042	28860	28427	20574
$\hat{\theta}_4$	20434	34572	26607	23567	25254	24819	17820
$\hat{\theta}_5$	18277	30259	23686	20932	22153	21739	15719
$\hat{\theta}_6$	15440	25899	20019	17785	19086	18706	13581
$\hat{\Theta}$	146912	246409	190862	168766	180252	177009	128349

Tāpat, kā gadījumā ar gamma sadalījumu, ir nepieciešama blīvumfunkcijas logaritmizēšana ar transformētu gamma funkciju:

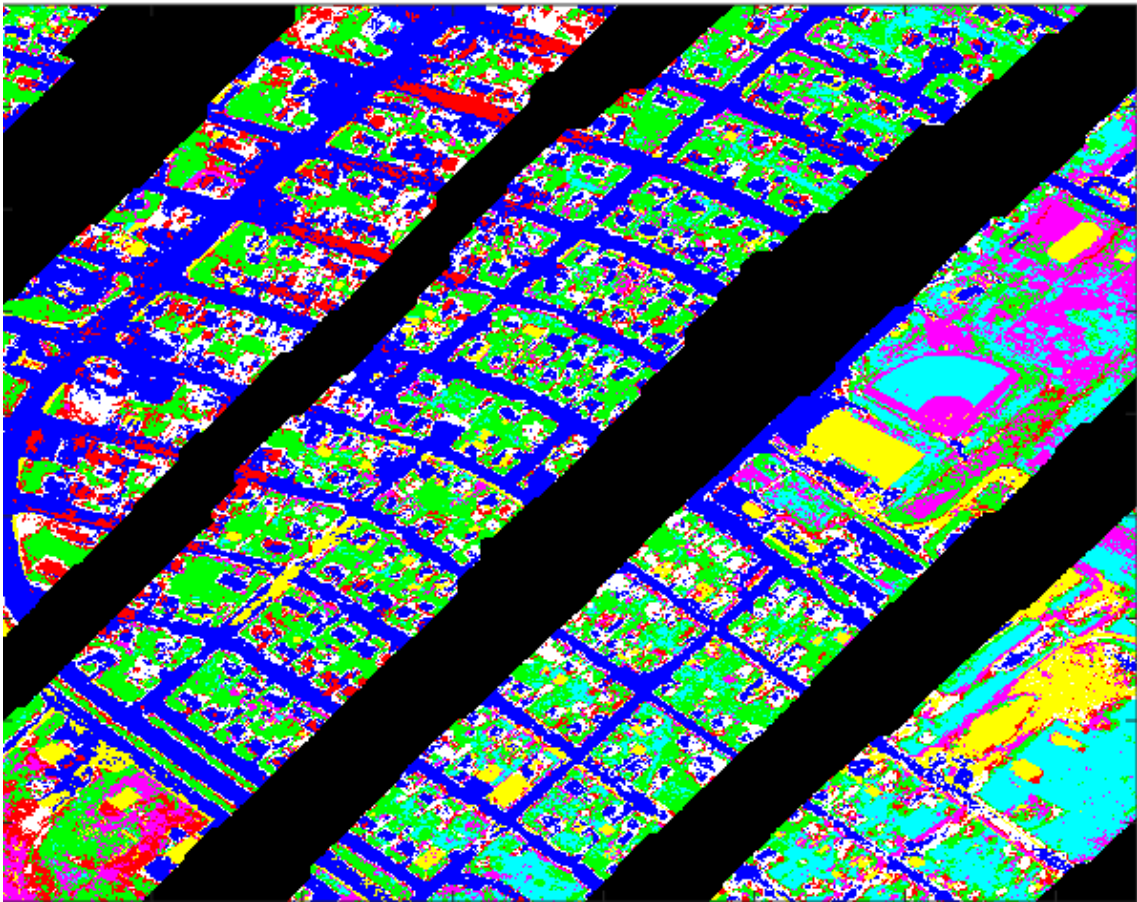
$$\begin{aligned}
& \ln(p_Z(z)) = \\
& = [\hat{\theta}] \ln \hat{\theta} \sum_{j=1}^{[\hat{\theta}]} \ln \left(1 - \frac{j}{\hat{\theta}}\right) + \ln \Gamma(\hat{\theta} - [\hat{\theta}]) \\
& \quad - \sum_{i=0}^n \left\{ [\hat{\theta}_i] \ln \hat{\theta}_i \sum_{j=1}^{[\hat{\theta}_i]} \ln \left(1 - \frac{j}{\hat{\theta}_i}\right) + \ln \Gamma(\hat{\theta}_i - [\hat{\theta}_i]) \right\} \\
& \quad + \sum_{i=1}^n (\hat{\theta}_i - 1) \ln(z_i) + (\hat{\theta}_0 - 1) \ln \left(1 - \sum_{i=1}^n z_i\right).
\end{aligned} \tag{2.17}$$



2.13. att. **Klasifikācijas rezultāti Dirihlē sadalījumam**

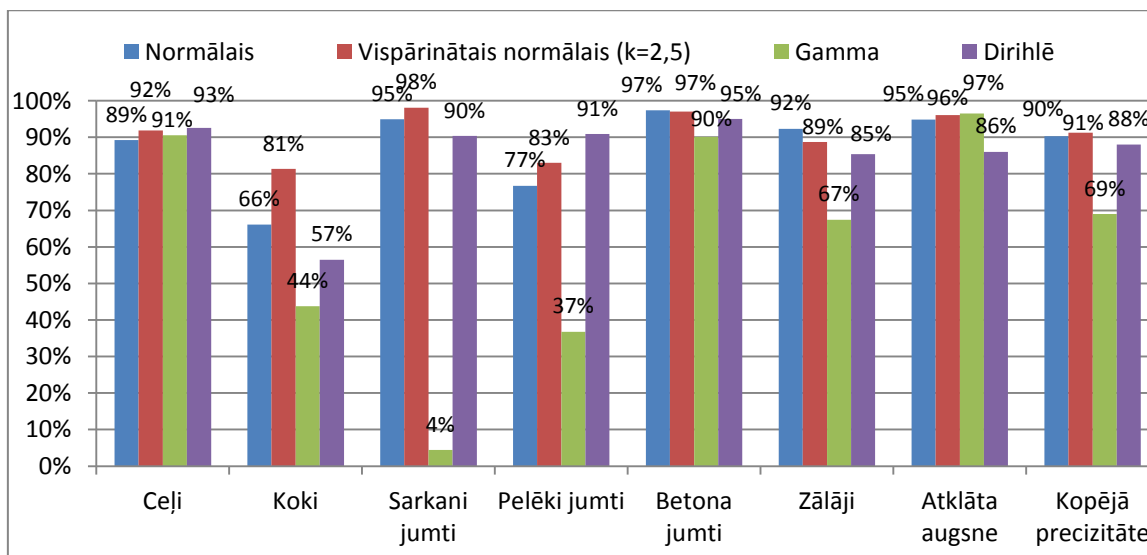


2.14. att. **Klasifikācijas kļūdas Dirihlė sadalījumam, klasifikators I**



2.15. att. **Klasifikācijas karte Dirihlė sadalījumam, klasifikators I**

Visbeidzot tiks apskatīts kopsavilkums visiem sadalījumu veidiem.

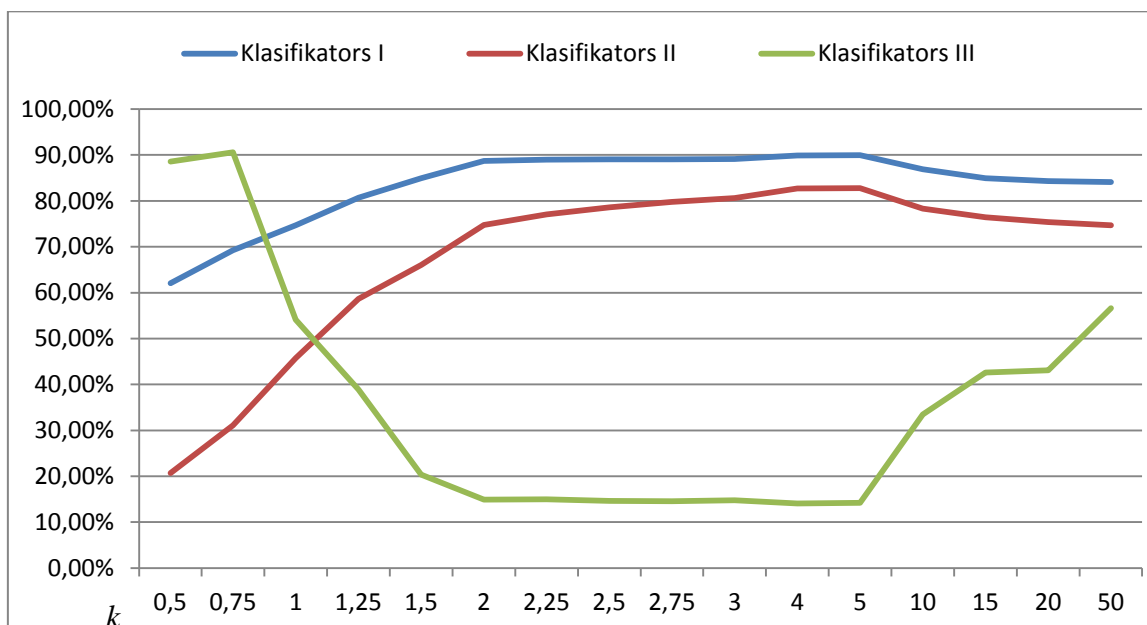


2.16. att. **Klasifikācijas rezultātu salīdzinājums visiem sadalījumiem**

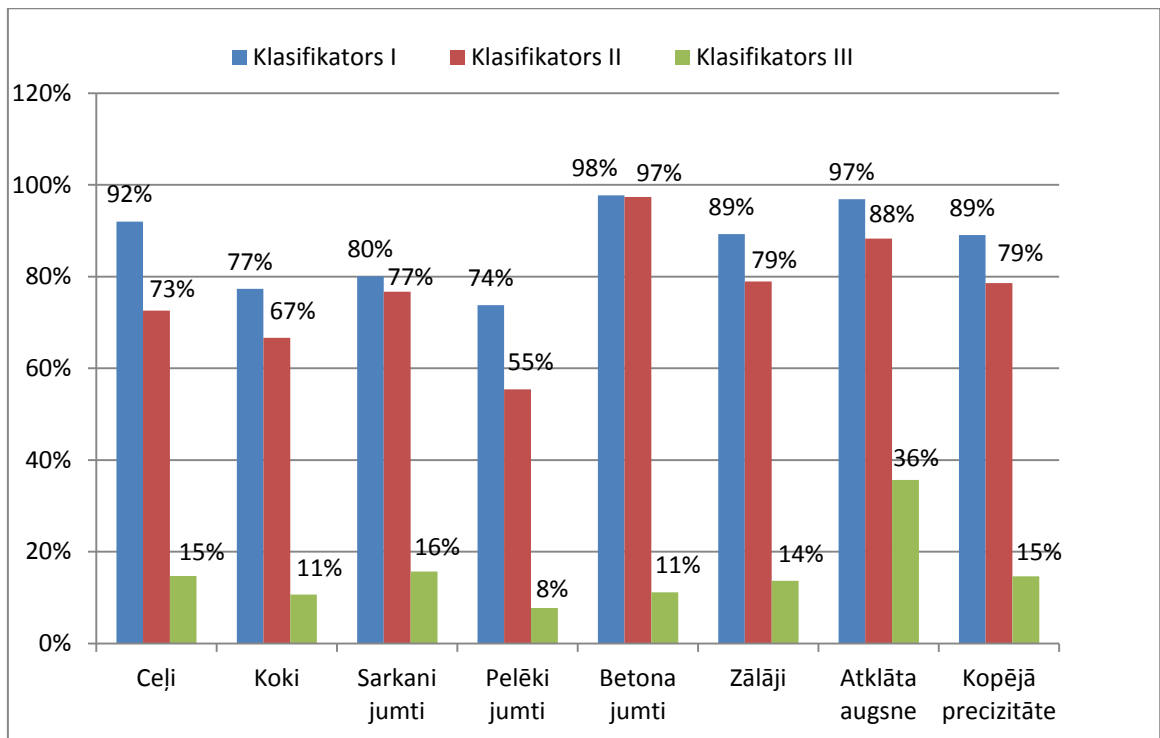
Redzams, ka visaugstākā kopējā precizitāte ir vispārinātajam normālajam sadalījumam ar $k = 2,5$, kas gan ir tikai pavisam nedaudz labāks rezultāts, kā normālajam sadalījumam. Vai šo rezultātu ir iespējams uzlabot?

2.6 Funkciju kombinācijas

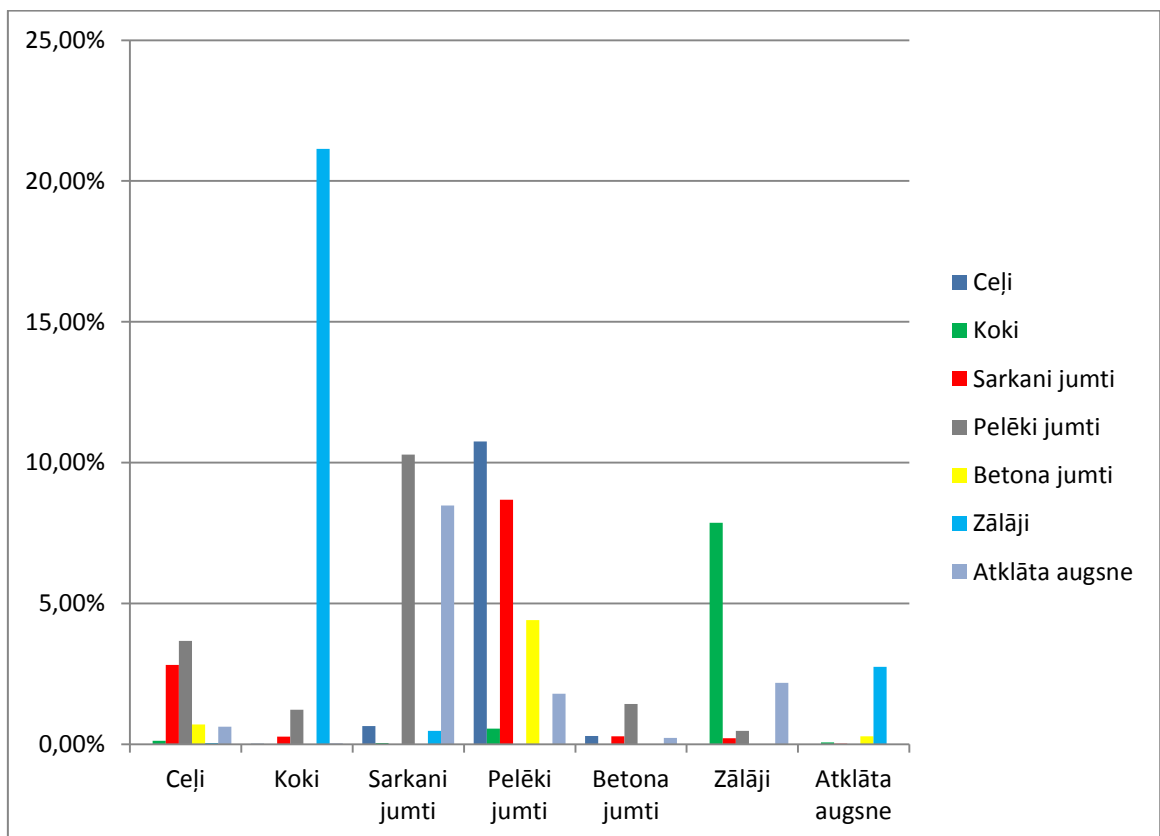
Noslēgumā tiks aplūkota klasifikācija, kas balstās uz divu sadalījuma blīvumfunkciju summu.



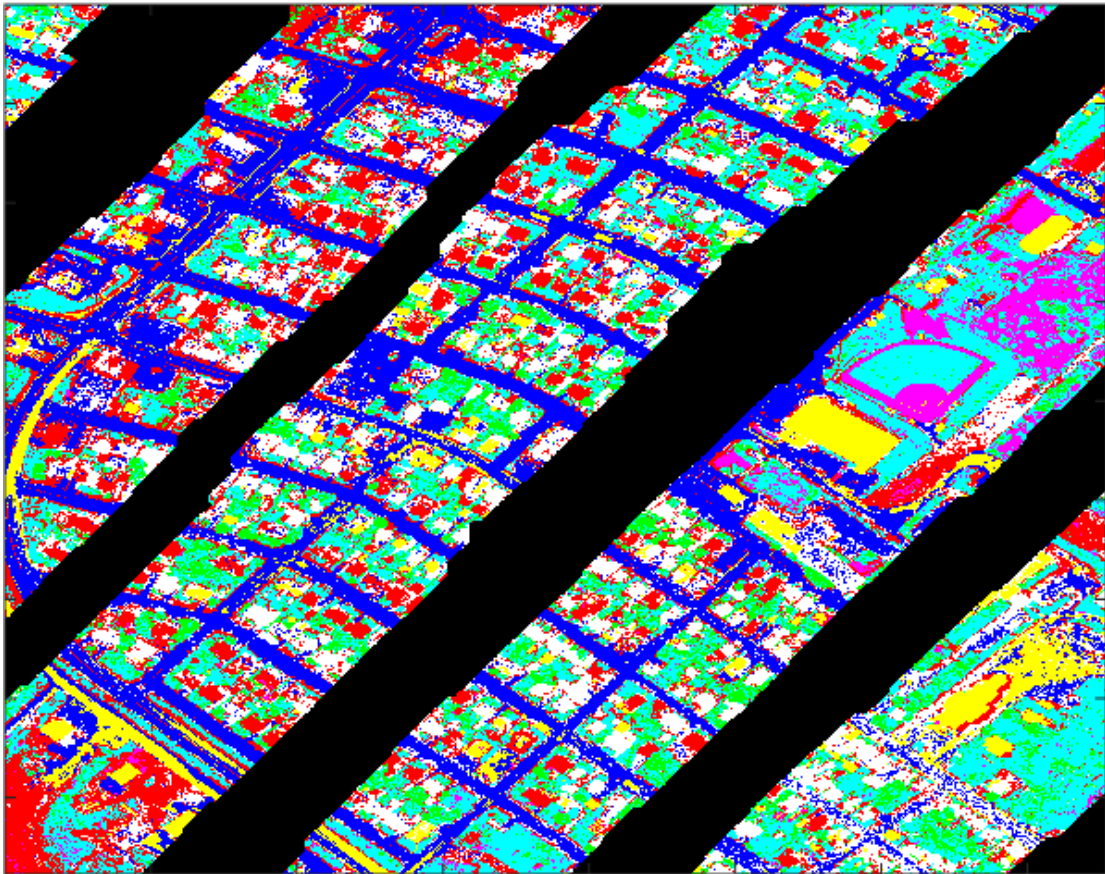
2.17. att. **Klasifikatoru kopējā precizitāte vispārinātā normālā sadalījuma kombinācijai ar gamma sadalījumu**



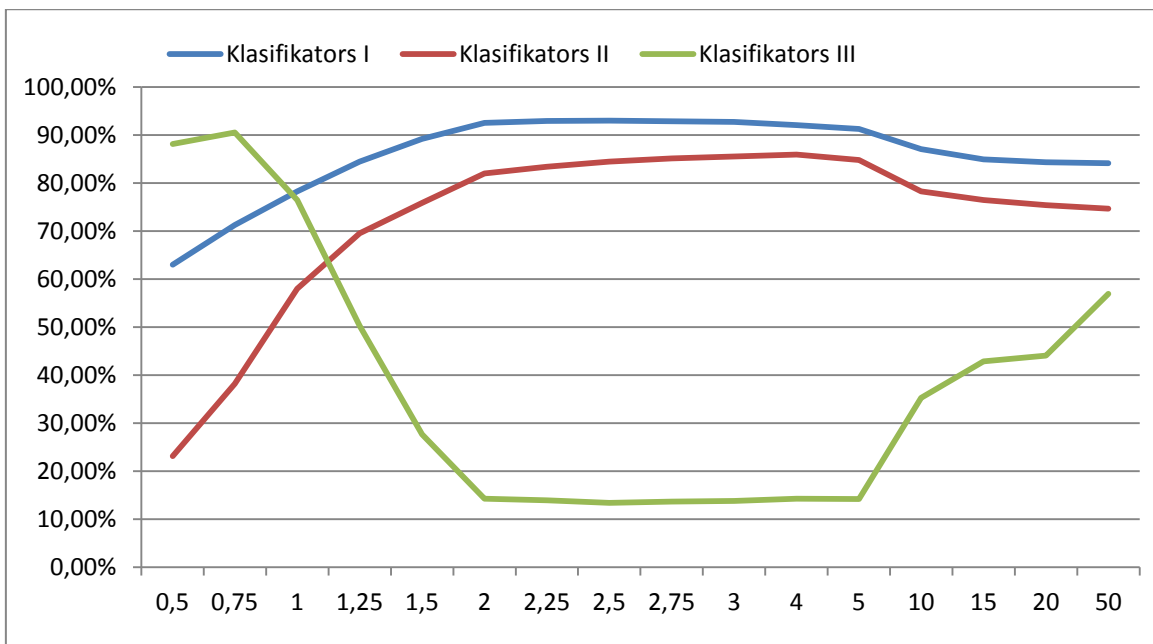
2.18. att. **Klasifikācijas rezultāti vispārinātā normālā sadalījuma (k=2,5) kombinācijai ar gamma sadalījumu**



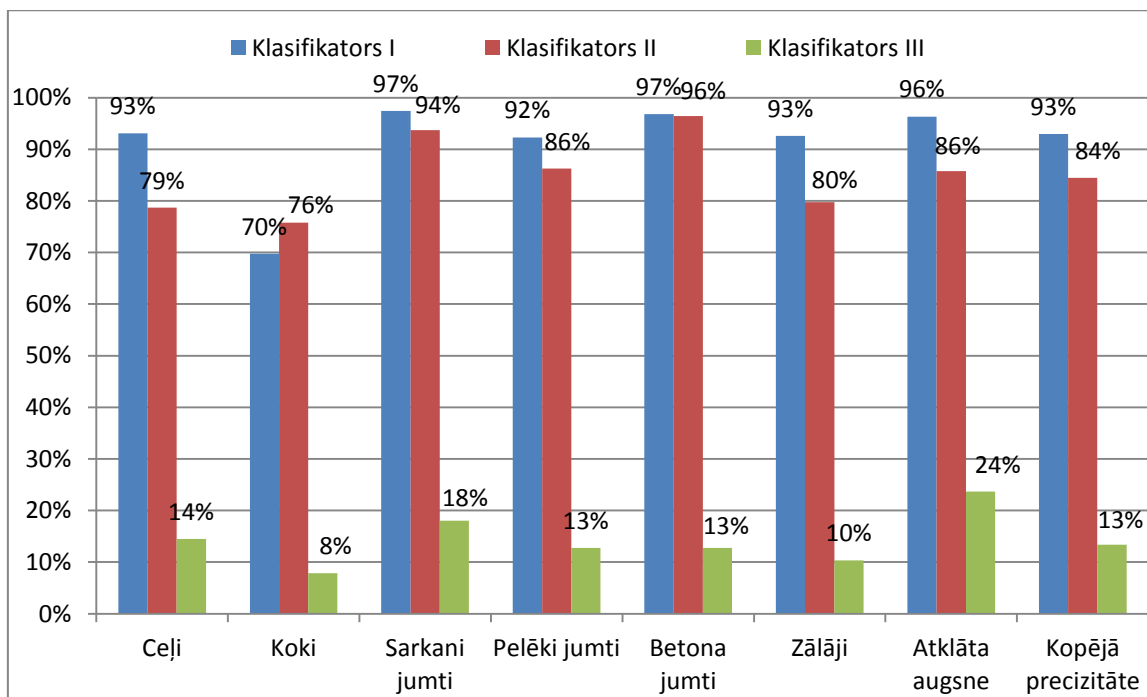
2.19. att. **Klasifikācijas kļūdas vispārinātā normālā sadalījuma (k=2,5) kombinācijai ar gamma sadalījumu, klasifikators I**



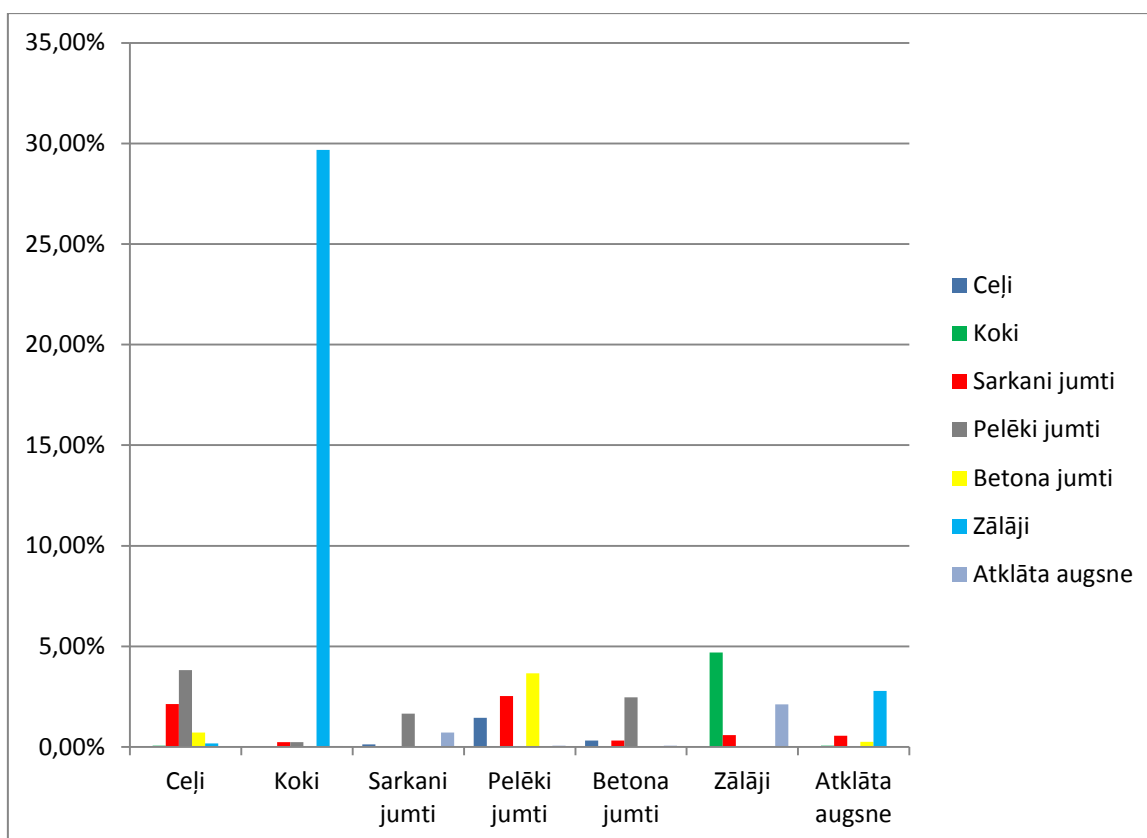
2.20. att. **Klasifikācijas karte vispārinātā normālā sadalījuma ($k=2,5$) kombinācijai ar gamma sadalījumu, klasifikators I**



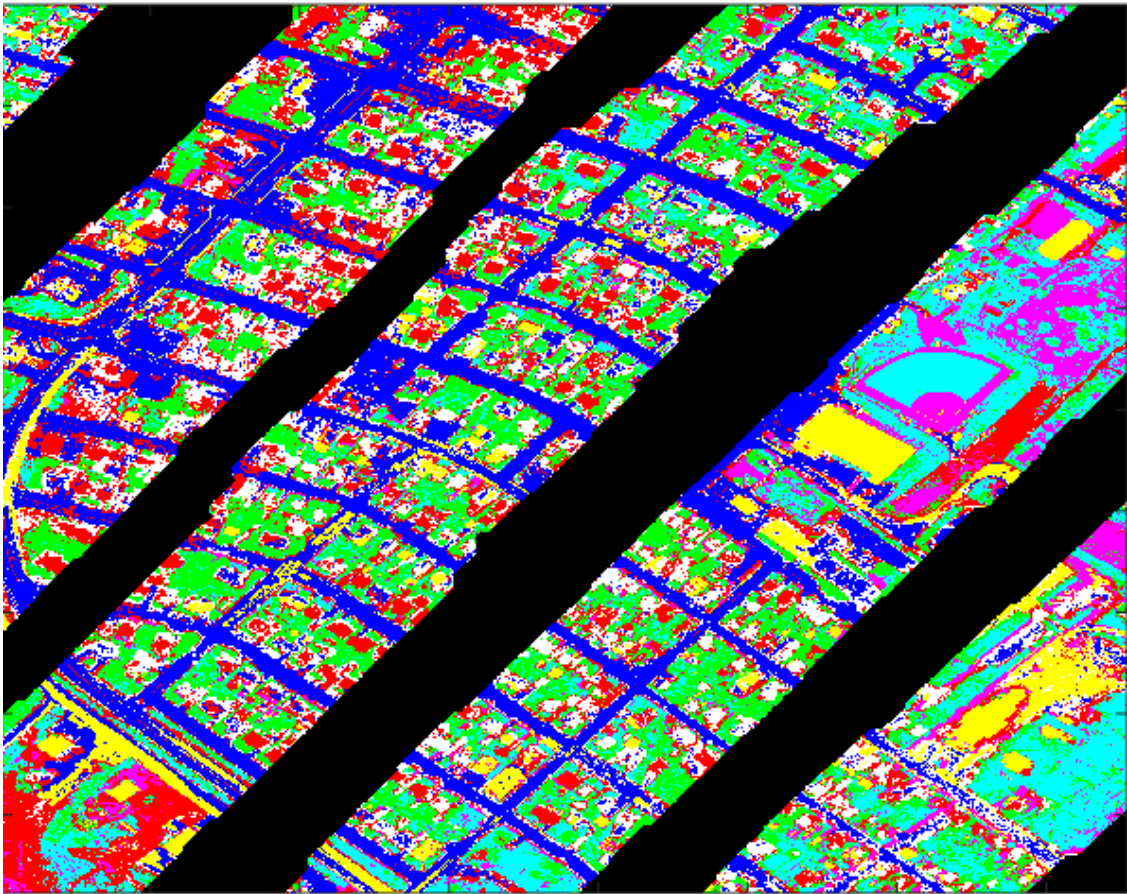
2.21. att. **Klasifikatoru kopējā precizitāte vispārinātā normālā sadalījuma kombinācijai ar Dirihlē sadalījumu**



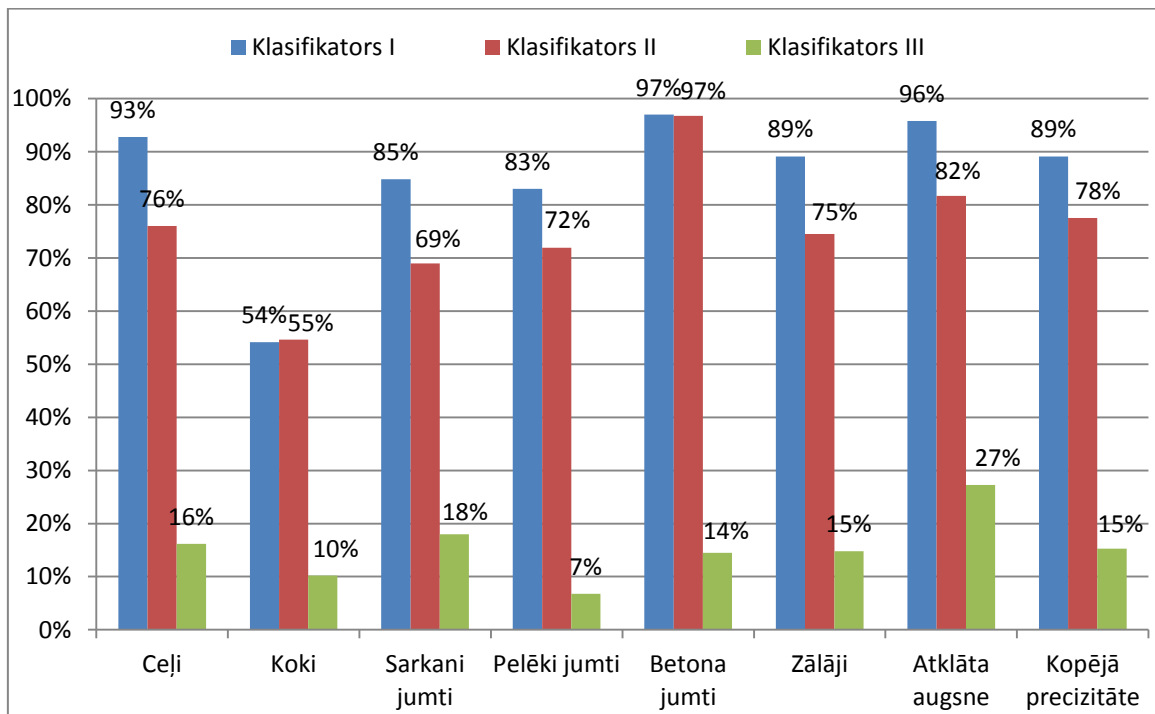
2.22. att. **Klasifikācijas rezultāti vispārinātā normālā sadalījuma ($k=2,5$) kombinācijai ar Dirihlē sadalījumu**



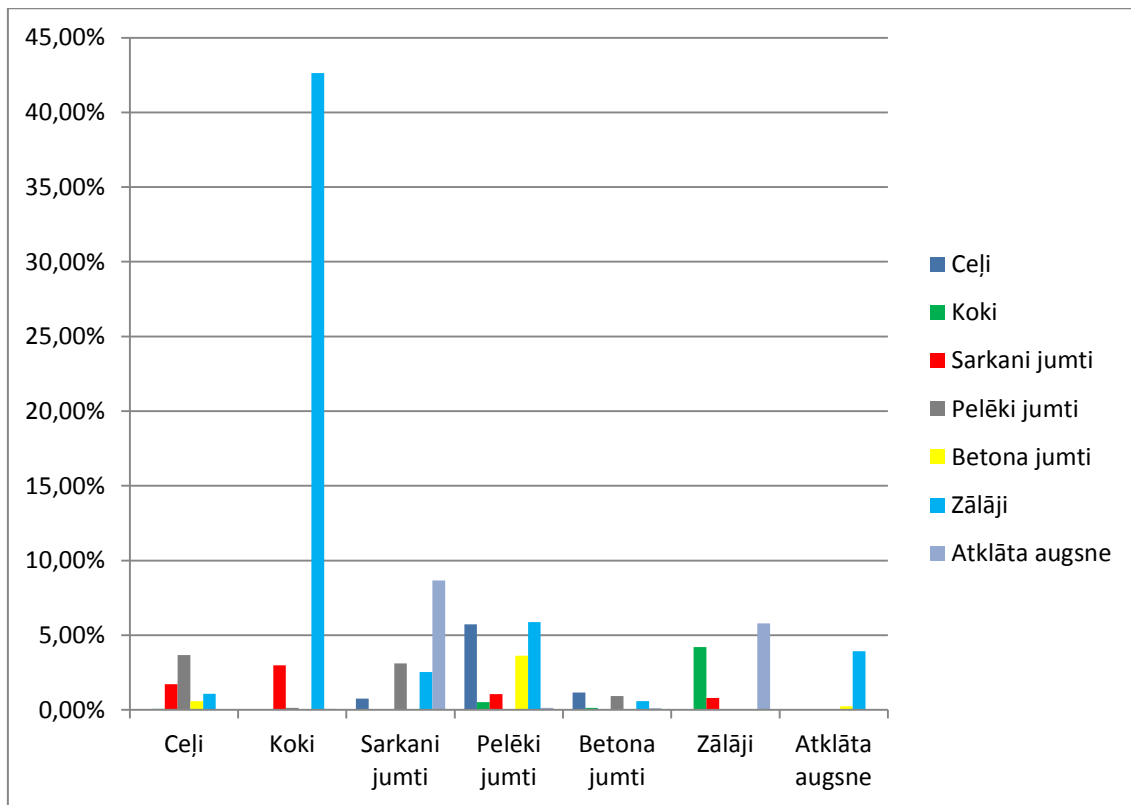
2.23. att. **Klasifikācijas kļūdas vispārinātā normālā sadalījuma ($k=2,5$) kombinācijai ar Dirihlē sadalījumu, klasifikators I**



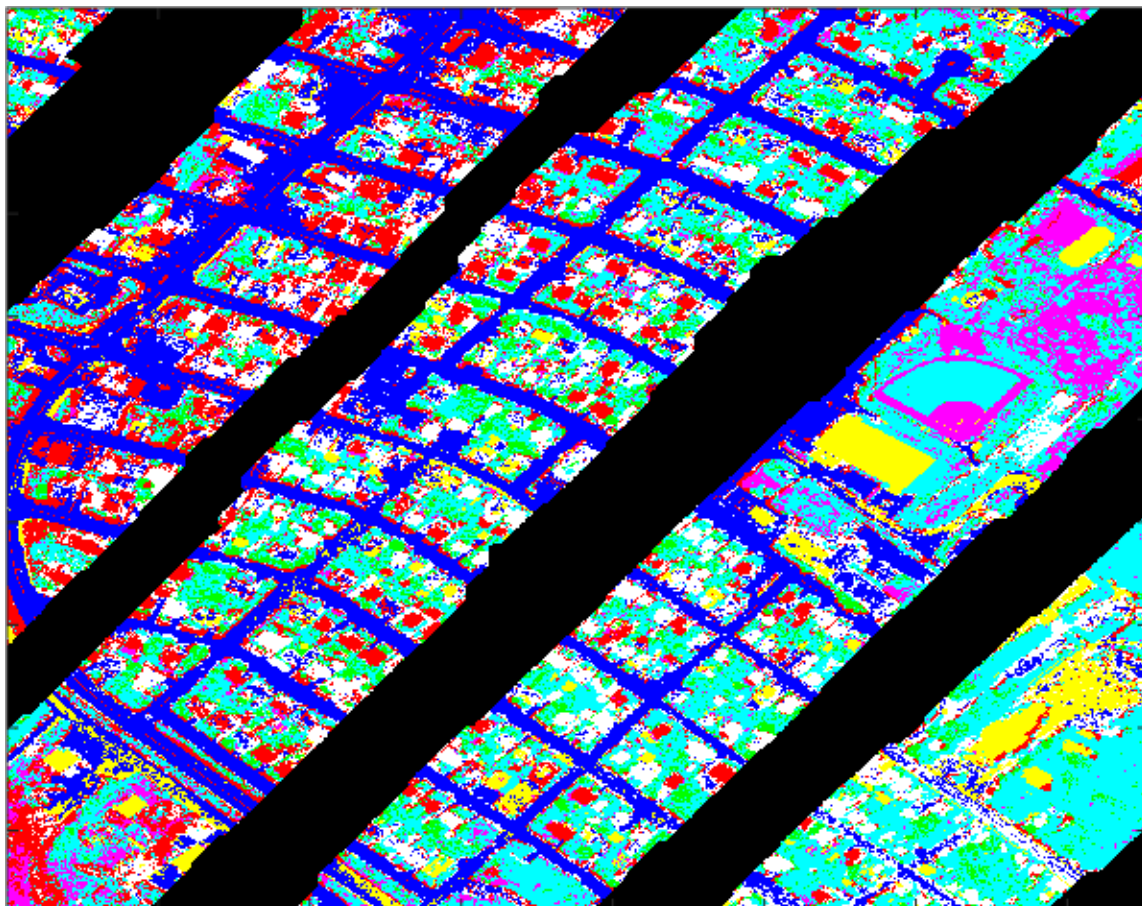
2.24. att. Klasifikācijas karte vispārinātā normālā sadalījuma ($k=2,5$) kombinācijai ar Dirihlē sadalījumu, klasifikators I



2.25. att. Klasifikācijas rezultāti Dirihlē sadalījuma kombinācijai ar Gamma sadalījumu

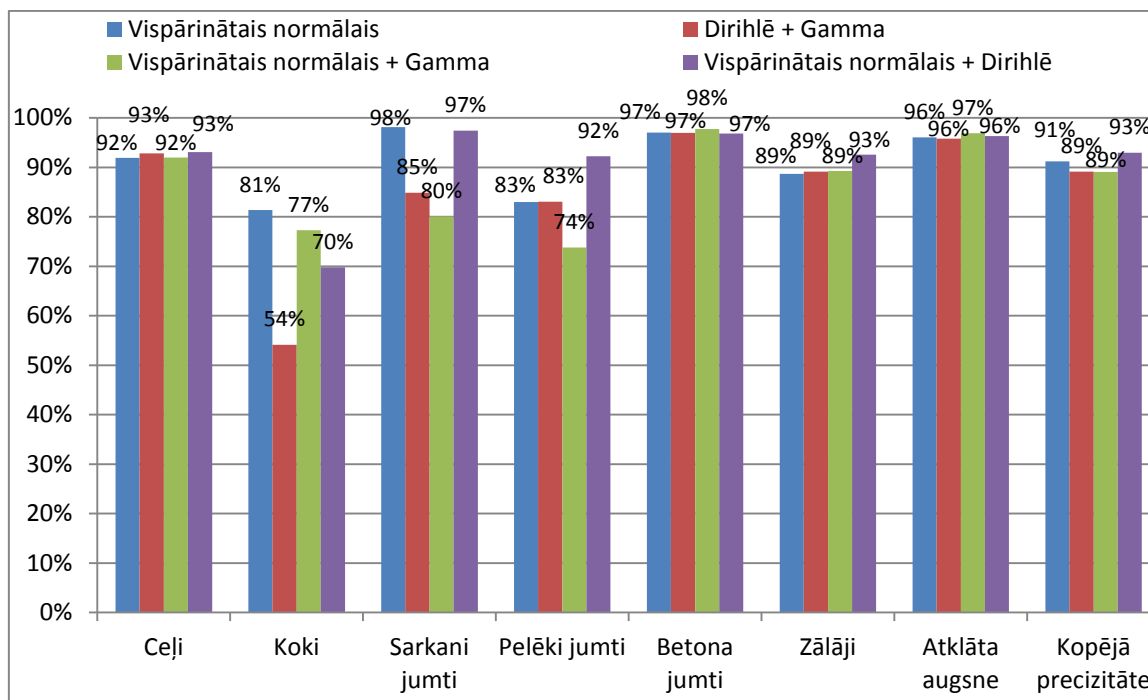


2.26. att. Klasifikācijas kļūdas Dirihlē sadalījuma kombinācijai ar Gamma sadalījumu



2.27. att. Klasifikācijas karte Dirihlē sadalījuma kombinācijai ar Gamma sadalījumu

Salīdzinot sadalījumu kombināciju rezultātus ar iepriekšējās sadaļas rezultātiem, redzams, ka visaugstākā kopējā precizitāte tiek iegūta kombinējot Dirihlē sadalījumu ar vispārināto normālo sadalījumu.



2.28. att. **Klasifikācijas rezultātu salīdzinājums vispārinātajam normālajam sadalījumam ar sadalījumu kombinācijām**

Secinājumi

Izpētes rezultātā ir konstatēts, ka arī Dirihlē un vispārinātais normālais sadalījums dod labus klasifikācijas rezultātus. Gamma sadalījums dod izcilus rezultātus atsevišķām klasēm, bet tā kopējā precizitāte ir zema. Vispārinātā normālā sadalījuma analīze rāda, ka visaugstākie rezultāti tiek gūti pie izliekumu raksturojošā parametra k vērtības 2,5.

Dažādu klasifikatoru analīzes rezultāti rāda, ka vizuālā un termālā attēla blīvumfunkciju summa konsekventi dod labākus rezultātus, kā citas klasificēšanas metodes.

Blīvumfunkciju kombināciju analīze deva visaugstākos precizitātes novērtējumus. Ir vērts izvest šajā virzienā padziļinātu analīzi, kā, piemēram, aplūkot klasifikatoru, kuram vizuālo un termālo attēlu apstrādā dažādu sadalījumu blīvumfunkcijas.

Izmantotā literatūra un avoti.

- [1] Dinuls R., Erins G, Lorencs A., Mednieks I., Member, IEEE, Siņica-Siņavskis J. *Tree Species Identification in Mixed Baltic Forest Using LiDAR and Multispectral Data* IEEE journal of selected topics in applied earth observations and remote sensing, vol. 5, No. 2, april 2012
- [2] Lorencs A., Mednieks I., Siņica-Siņavskis J. *Classifications of Multisensoral Images With Different Spatial Resolution* Elektronika ir Elektrotehnika
- [3] 2014 IEEE GRSS Data Fusion Contest. <http://www.grss-ieee.org/community/technical-committees/data-fusion/>
- [4] Kotz S., Balakrishnan N., Johnson N. L. (2000) *Continuous Multivariate Distributions* Wiley-Interscience Publication
- [5] Solaro N. (2004) *Random variate generation from Multivariate Exponential Power distribution* Statistica & Applicazioni Vol. II, n. 2, 2004

Maģistra darbs „Daudzdimensionālu sadalījumu lietošanas aspekti multispektrālu attēlu analīzē” izstrādāts LU Fizikas un matemātikas fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Mārtiņš Puķītis _____

Rekomendēju darbu aizstāvēšanai

Vadītājs: prof. Dr.habil. math. Aivars Lorencs _____02.06.2015

Recenzents: _____

Darbs iesniegts matemātikas nodaļā __.06.2015.

Metodiķe: Dzintra Holsta _____

Darbs aizstāvēts maģistra gala pārbaudījuma komisijas sēdē

____ 06.2015. prot. Nr. _____

Komisijas sekretāre: lektore Inese Bērziņa