

Rīgas Klasiskā ģimnāzija

**Attēlu atpazīšanas algoritmu un datormetožu izmantošana  
astronomisko objektu klasifikācijā**

Zinātniski pētnieciskais darbs informātikā

Darba autors:

Aleksejs Sazonovs

11.klase

Darba vadītājs:

Rīgas Klasiskās ģimnāzijas informātikas skolotājs,

M.Ed. Sergejs Bratarčuks

Rīga, 2010.

# Satura rādītājs

levads.....	2
1. Teorētiskā daļa .....	5
1.2. Galaktiku klasifikācijas zinātniskā nozīme .....	5
1.3. Attēlu klasificēšanas algoritmi .....	6
1.3.1. Prasības metodei astronomijas uzdevuma kontekstā.....	6
1.3.2. Attēlu atpazīšanas algoritmi .....	6
1.3.3. ET-DIC metode.....	8
1.3.4. Datorredze.....	8
1.3.5. Statistisko datu par galaktikām iegūšana ar apakšlogu metodi .....	9
1.3.6. Risinājumu koki.....	9
1.3.7. Ansambļa metode .....	10
1.4. Nodaļas apkopojums .....	10
2. Praktiskā daļa .....	11
2.1. Attēlu atpazīšanas algoritma testēšana.....	11
2.1.1. Eksperimenta nosacījumi .....	11
2.1.2. Sistēmas darba novērtējums .....	12
2.1.3. Atpazīšanas sistēmas uzstādījumu parametru noteikšana .....	13
2.1.4. Testēšanas rezultāti.....	17
2.2. Rezultātu publikācija on-line datubāzē.....	18
Rezultāti.....	19
Secinājumi.....	20
Noslēgums .....	21
Izmantotas literatūras saraksts .....	22
Anotācija.....	23
Pielikumi .....	24

## levads

Astronomija ir viena no senākajām zinātnēm. Tā radās nepieciešamības dēļ orientēties laikā un telpā. Reizē ar to debesu vērošana veidoja cilvēku kā domājošu būtni. Ir zināmas liecības par cilvēku sistemātisku nodarbošanos ar astronomiju vēl neolīta laikmetā. Viens no visplašāk pazīstamiem skatu laukumiem, kas ir orientēts pēc debess pusēm, ir Stounhedža. Līdzīgi laukumi, kas acīmredzot, tika izmantoti kulta astronomijā, tika atrasti arī citās vietās, tai skaitā arī Baltijas valstu teritorijā (Eremeeva A.I., Cicins F.A., 1989, lpp. 18,19).

Astronomija kā zinātne mūsdienu izpratnē attīstījās tur, kur radās rakstniecība. Rakstniecība ļāva drošāk uzglabāt un sistematizēt no novērojumiem iegūtās ziņas. Tas sekmēja sarežģītāku un dziļāku likumsakarību atklāšanu, apvienojot attālu paaudžu pūliņus (Eremeeva A.I., Cicins F.A., 1989, lpp. 25).

Piemēram, babilonieši atstāja praktiski pilnu sarakstu ar aptumsumiem sākot ar 763. gadu p.m.ē., kura daļu izmantoja Ptomelejs, lai uzlabotu teoriju par mēness kustību (Eremeeva A.I., Cicins F.A., 1989, lpp. 29).

Jau VI gadsimtā p.m.ē. ķīnieši izdalīja (galaktiku) Piena ceļš kā nepazīstamu dabas parādību (Eremeeva A.I., Cicins F.A., 1989, lpp. 37). Teorētiskā astronomija radās daudzās civilizācijās – senajā Ķīnā, Indijā, Grieķijā un Mezopotāmijā.

Līdz ar materiālistiskiem uzskatiem par debesu ķermeņu kustību radās arī samērā oriģinālas, bet mistiskas teorijas. Piemēram, Pitagora skaitļu teorija. Viņa sekotāji centās atrast debesu ķermeņu kustībā „parasto skaitļu maģiju”. Šīs abstraktās mistikas dēļ viņi saskārās ar nepārvaramiem un neatbildētiem jautājumiem, bet no otras puses tā bija pirmā dabaszinātņu „matemizācija” (Eremeeva A.I., Cicins F.A., 1989, lpp. 63). Pitagora sekotāju idejas vēlāk palīdzēja Kepleram. Viņš skeptiski attiecās pret ideju „aprakstīt pasauli ar parastiem skaitļiem”, bet viņš varēja racionāli izmantot datus, kurus uzkrāja pitagorieši.

Astronomija kā zinātne attīstījās pateicoties novērošanas instrumentu pilnveidošanai, bet, kas ir ne mazāk svarīgi, – arī pateicoties datu sistematizēšanai. Mirzo Ulugbeka novērojumu dati nonāca Tiho Brage rīcībā, kurus no viņa tad arī „mantoja” Keplers, kurš radīja debesu mehānikas likumus (Krieg, 2005, lpp. 150-166). Šie likumi joprojām ir mūsdienu astronomijas un kosmonautikas „ābece”

Mūsdienās cilvēks var redzēt un pētīt neticami tālus Visuma nostūrus. Astronomiskās informācijas krājumi ir ļoti lieli. To apstrāde un analizēšana, izmantojot pat visjaudīgākos datorus, ir

neticami sarežģīts un bieži vien neizpildāms uzdevums. Šeit palīgā nāk datortīkli, mūsdienu datoru un matemātiskās metodes.

Kā šādu mēģinājumu var minēt projektu SETI (ārpus zemes saprāta meklēšana), kura ietvaros zinātnieki cenšas izpildīt ambiciozu un fantastisku uzdevumu – atrast saprāta liecības Visumā. Aresibo izvietotais teleskops reālā laika režīmā „raksta” milzīgus datu masīvus, kas tiek sadalīti fragmentos un izsūtīti uz projekta dalībnieku datoriem visā pasaulē. Visuma radiotrokšņu haosā notiek (pagaidām neveiksmīgi) mēģinājumi atrast likumsakarības no kosmosa nākošiem signāliem.

Tomēr astronomijas „ambīcijas” nav ierobežotas tikai ar SETI projektiem. Uzdevumu ir daudz, un tiem ir liela zinātniskā nozīme. Piem.:

- Kad notiek uzplaisnījums spektrā vai attēlā – vai tas ir reāls signāls, vai troksnis?
- Kad signāls ir pastāvīgs, nevis mainīgs?
- Kā pareizi izprast telpā notiekošos procesus, zvaigžņu izvietojumu galaktikā Visumā, kura dimensijas ir vairāk kā trīs?.. Un daudzus citus (Eberly College of Science, 2009).

Viena no informātikas viedokļa interesantākajām astronomijas problēmām ir galaktiku tipu automātiska klasifikācija. Klasifikācija fundamentālā zinātnē bieži vien palīdz saprast procesu iemeslus. Bioloģijā tās ir Karla Linneja un Darvina sistematizācijas, kuras palīdzēja saprast dzīvo organismu attīstības likumus. Ķīmijā ir elementu periodiskā sistēma, kuras pamatojoties uz elementu grupēšanu, izskaidro elementu īpašības. Fundamentālajā (kas nav orientēta uz šī brīža inženieruzdevumu risināšanu) astronomijā galaktiku klasifikācija ir ļoti svarīga. Tā ļauj spriest par vienu vai otru „zvaigžņu debesu” gabalu vecumu, izskaidrot Visuma attīstību, veikt reakcionāru analīzi – saprast, kā attīstījās Visums pēc Lielā sprādziena.

Šīs zinātniskās problēmas aktualitāti nosaka tas, ka pēdējos gados astronomisku datu daudzums aug ģeometriskā progresijā. Virszemes teleskopi, orbitālais teleskops Habbls nodrošina zinātniekus ar milzīgiem Visuma attēlu masīviem. Objektu daudzums optiskajā, radio un rentgena spektrā ir skaitāms miljardos. Bez šaubām, nepietiks ar visu Zemes zinātnieku spēkiem, ja ar sistematizācijas jautājumiem nodarbosies vienīgi cilvēks.

Palīgā astronomijai var nākt informācijas tehnoloģijas un datoru metodikas, kas tiek veiksmīgi izmantotas citās sfērās – bioloģijā un inženierzinātnēs. Ņemot vērā mūsdienu datoru veikspēju, to atmiņas apjomu, liela apjoma grafisko datu analīzes uzdevums ir izpildāms. Izmantojot datoru

sistēmas objektu atpazīšanai astronomijā, var panākt progresuastronomiskās klasifikācijas uzdevumos. Dotajam pētījumam ir lietišķs raksturs, taču tā rezultāti ir noderīgi, pirmkārt, fundamentālajā astronomijā, kā arī tām zinātnes nozarēm, kurās ir nepieciešama automātiska liela apjoma attēlu apstrāde.

PĒTĪJUMA OBJEKTS: Statistiskās astronomijas datormetodes.

PĒTĪJUMA PRIEKŠMETS: Objektu atpazīšanas algoritmi, kas orientēti uz galaktiku klasifikācijas uzdevuma atrisināšanu.

PĒTĪJUMA HIPOTĒZE: Ir iespējama attēlu atpazīšanas datorsistēmas parametru pielāgošana un adaptācija, kas ļautu veikt un sasniegt automātisko galaktiku klasificēšanu ar vidējo precizitāti ne mazāk kā cilvēkam.

PĒTĪJUMA MĒRĶIS: Paplašināt bioinformātikā izmantojamās datoru sistēmas lietošanas iespējas attēlu atpazīšanā, adaptējot tās astronomijas zinātniskiem uzdevumiem.

PĒTĪJUMA UZDEVUMI:

- Izanalizēt pieejamo literatūru un interneta avotus pēc izvēlētā temata.
- Noteikt informācijas metožu pieejamību izmantošanai mūsdienu astronomijas uzdevumos.
- Noteikt automātisko attēlu atpazīšanas sistēmu rašanās un attīstības pamatetapus. Sniegt pārskatu par mūsdienu attēlu atpazīšanas metodēm to izmantošanu kontekstā izvēlētajai astronomijas zinātnes problēmai.
- Izvēlēties eksperimentālajam pētījuma programmu paketi, adaptēt tās uzstādījumus, lai varētu lietot galaktiku klasificēšanas problēmas risināšanai. Noteikt vissvarīgākos parametrus un mainīgos, kas nepieciešami klasificēšanai un kurus lietojot tiks iegūti vislabākie rezultāti.
- Noteikt sistēmas darba precizitātes minimālo un vēlamu kritēriju novērtēšanu.
- Veikt izvēlētās galaktiku attēlu kopas automātisko klasifikāciju. Iegūt eksperta novērtējumu saņemto datu relevantumam.
- Iegūtos rezultātus publicēt, lai tie būtu pieejami turpmākai izmantošanai astronomiskajai savienībai.

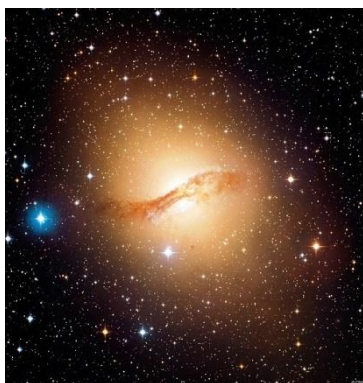
## 1. Teorētiskā daļa

### 1.1. Galaktiku klasifikācijas problēma

Informātika tiek plaši izmantota lielākajā daļā astronomisko pētījumu. Milzīgs datu apjoms un izskaitļošanas sarežģītums padara lietišķo informātiku par „vistuvāko” nozari astronomijai.

Daži no esošajiem projektiem fokusējas uz astronomisko attēlu apstrādi. Grūtības apstrādē rodas tāpēc, ka katru gadu brīvajā pieejā parādās ap 35 TB jaunu datu. Organizācija ST-ECF un Canadian Astronomy Data Centre (CADC) ir apvienojuši savus spēkus jaunas paaudzes tīkla arhīva izveidē. Viņu sistēma atvieglo nepieciešamo attēlu iegūšanu un to apstrādi.

Galaktikas – vislielākās Visuma daļas. Mūsu galaktika Piena ceļš sastāv no  $10^{11}$  zvaigznēm, turklāt tā nav lielākā no novērotajām (Elske V.P. Smith, 1987, lpp. 378). Pagājušā gadsimta sākumā Edvīns Habbls sadalīja galaktikas divās lielās kategorijās. Pastāv galaktikas, kuras pēc formas ir līdzīgas regbija bumbai (att.1.) un kuras tiek sauktas par elipsveidīgām un kuras ir līdzīgas spirālēm – spirālveidīgās (William J. Kaufmann, 1977, lpp. 365-366). Uz dažām galaktikām mēs, Zemes iedzīvotāji, raugāties no sāna. Zinātnieki uzskata, ka tās ir parastās spirālveidagalaktikas, kas līdzinās mūsu galaktikai. Šī klasifikācija nav pilnīgākā, tomēr vairumā gadījumu tā darbojas.



1. attēls „Galaktiku veidi”

### 1.2. Galaktiku klasifikācijas zinātniskā nozīme

Teorētiskajā modelēšanā astronomi noskaidroja, ka spirālveida galaktiku saplūšanas rezultātā var rasties spirālveidīgie. Tālāko zvaigžņu un kosmisko gāzu saplūšanas procesā elipsveidīgie var kļūt par spirālveidīgiem. Reālo galaktiku attēli, kas uzņemti ar vismodernāko teleskopu palīdzību, parāda, ka šie procesi patiešām norisinās. Cik bieži notiek saplūšanas? Vai tās ir svarīgas? Vai katra galaktika

savas pastāvēšanas laikā iziet cauri šādām transformācijām? Lai to apstiprinātu, ir nepieciešams liels datu apjoms par galaktiku tipiem. Projekts Galaxy Zoo (brīvprātīgo galaktiku, kas iegūtas ar teleskopa Sloan Digital Sky Survey palīdzību, klasifikācijas programma) palīdzēja spert pirmo soli nepieciešamo datu saņemšanā, tomēr ļoti garš ceļš vēl ejams.

Šo datu pieprasījums pierāda automātiskās morfoloģiskās galaktiku klasifikācijas nepieciešamību.

### **1.3. Attēlu klasificēšanas algoritmi**

#### **1.3.1. Prasības metodei astronomijas uzdevuma kontekstā**

Avotu apskats un konsultācijas ar dažu astronomisko projektu darbiniekiem norādīja uz to, ka šodien vēl nepastāv gatavi programmu produkti, kas spētu risināt uzstādīto galaktiku klasifikācijas uzdevumu. Tomēr pastāv pietiekami daudz objektu atpazīšanas metožu. Konkrētas metodes izmantojums ir atkarīgs no konkrēta uzdevuma.

Lai atrisinātu galaktiku klasifikācijas uzdevumu, ir nepieciešams, lai izvēlētais algoritms būtu:

- izmantojams jaunam un sev nestandarta uzdevumam;
- spējīgs atpazīt dažādu klašu objektus;
- spējīgs strādāt ar uzņemtajiem (nofotografētajiem) attēliem dažādos apstākļos;
- spējīgs ņemt vērā apgaismošanas, mēroga, skatupunkta, objekta orientācijas izmaiņas trijās dimensijās;
- noturīgs pret vizuāliem kropļojumiem un „trokšņainu” fonu.

Šīs prasības izriet no astronomiskās fotografēšanas īpatnībām, tipiskas galaktiku formas un iespējas, ka galaktika var atrasties pret vērotāju ar „jebkuru pusi”.

#### **1.3.2. Attēlu atpazīšanas algoritmi**

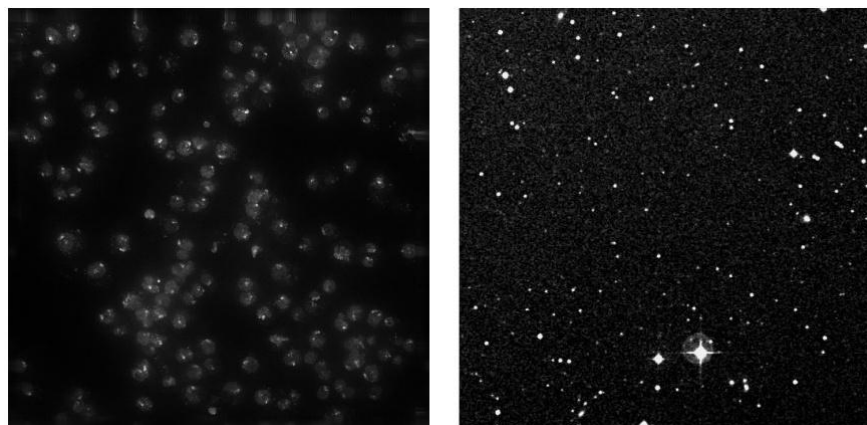
Literatūrā (Anil K. Jain, 2000, lpp. 6) ir aprakstītas dažas principiāli atšķirīgas metodes un pieejas objektu atpazīšanai (attēli, skaņu faili utt.). Katra metode vai metožu grupa var būt veiksmīgāk izmantota noteiktai uzdevumu klasei. Metodes ir apkopotas 1. tabulā.

Atpazīšanas metode	Datu attēlošanas veids	Atpazīšanas funkcijas	Novērtēšanas kritērijs
„Tēlu pretstatīšana”	Paraugi, pikseli, līknes	Korelācija, distances noteikšana	Klasifikācijas kļūda
„Statistiskā”	Elementu īpašības	Diskriminanta funkcija	Klasifikācijas kļūda
„Sintaksiskais vai struktūras”	Vienkāršākie objekti	Noteikumi, gramatika	Uztveršanas kļūda
„Neironu tīkli”	Paraugi, pikseli, elementu īpašības	Tīkla funkcija	Vidējā kvadrātiskā kļūda

Bez metožu apskata pētījuma ietvaros tika veikta University of Liège darbinieka Dr. Raphaël Marée intervija (4. pielikums). Viņš izteica pieļāvumu, ka izvirzītajām prasībām atbilst GIGA Bioinformatics grupas no University of Liège „CVPR05” (Raphaël Marée, 2005) konferencē piedāvātā metode. Šī metode attiecas pie „Statistisko” metožu grupas.

Šī metode daļēji tika izmantota vienā no pētījumiem (Raphaël Marée, Pierre Geurts, Louis Wehenkel, 2006). Ar tās palīdzību notika automātiska bioloģisku attēlu klasificēšana.

No algoritma viedokļa – nav jābūt principiālai atšķirībai starp apstrādājamiem attēliem. Var veikt visvienkāršāko eksperimentu – blakus drozofilas šūnas fotogrāfijai novietot zvaigžņotas debess attēlu (att. 2).



2. attēls. Pa kreisi – drozofilas šūnas, pa labi – zvaigžņotās debess apgabals



Izskatās, ka starp galaktikām un mikroskopiskiem dzīvo organismu elementiem nav nekādas tiešas saiknes. Tomēr vizuāli tie ir ļoti līdzīgi. Tas ļauj izteikt pieļāvumu, ka bioloģisku un kosmisku objektu atpazīšanas algoritmi var būt līdzīgi.

### **1.3.3. ET-DIC metode**

Zinātnieku grupa „Bioinformatics” no University of Liège konferencē, (CVPR 2005), parādīja ET-DIC attēlu klasificēšanas metodi. Šajā metodē attēls tiek sadalīts daudzos nejauši iegūtos „apakšlogos” (subwindow) un apraksta tos, izmantojot daudzdimensiju vektorus, kas sastāv no pikseļu virknes (2. pielikums). Pēc tam metode izmanto ārkārtīgi randomizēto risinājumu koku (ensemble of extremely randomized decision trees) kopu klasifikācijas modeļa izveidošanā. Jauna attēla klases prognozēšanai metode apkopo risinājumu koku sastādītās prognozes un izmanto visbiežāk atkārtoto prognozi lēmuma pieņemšanai.

Avotos ir minēts, ka šī metode ir parādījusi augstu efektivitāti uz dažādām attēlu bāzēm, tādām kā skaitļi, sejas, priekšmeti, ēkas, fotogrāfijas. Viens no visgrandiozākajiem pētījumiem ar šīs metodes izmantošanu bija par bāzi ņemot 10000 rentgena uzņēmumu. Iegūtie rezultāti bija tuvi etalonam<sup>1</sup>.

Dotajā pētījumā šī metode tika izmantota astronomijas uzdevumiem.

### **1.3.4. Datorredze**

Spējas mācīties piemīt ikvienai ar intelektu apveltītai. Herberts Simons (Herbert Simon) nodefinēja mācīšanu šādi: „*Mācīšana ir jebkuras izmaiņas sistēmā, kas uzlabo uzdevuma atrisināšanu tam atkārtojoties, vai cita uzdevuma, kura pamatā ir tādi paši dati, atrisināšanu*” (Simon, 1983).

Mašīnāmācība pa lielam ir lietderīgas informācijas meklēšana iespējamo jēdzienu telpā un to korekts apkopojums (Luger, 2001, lpp. 373). Vienas no šādiem mācību veidiem ir mācīšanās ar skolotāju (supervised learning).

Šīs klases algoritmi pieņem, ka mācību dati ir klasificēti. Mācāmajai sistēmai tiek paziņots, pie kuras klases attiecas dotais piemērs (Luger, 2001, lpp. 373).

Pirmajā posmā tiek izveidota datu bāze (galaktiku attēlu bāze). Tā tiek sadalīta divās daļās: attēli treniņam un attēli pārbaudei.

---

<sup>1</sup>Etalona rezultāti – labākie iegūtie rezultāti no vienas automātiskās attēlu klasificēšanas sistēmas.

Ar algoritma palīdzību tiek izdalītas attēlam piemītošās raksturiezīmes (features). Pamatojoties uz tām, tiek būvēts modelis. Turpmāk modelis tiek izmantots klasifikācijai.

Lai noteiktu modeļa kvalitāti, atlasītie attēli tiek nosūtīti klasifikatoram. Ar iepriekš izveidotā modeļa palīdzību klasifikators apstrādā attēlus, piešķirot tiem klasi. Pēc tam tiek izskaitļots „Error rate” — sistēmas kļūdainas izstrādes procents. Ja iegūtais rezultāts nav apmierinošs, skolotājs turpina sistēmas trenēšanu.

### **1.3.5. Statistisko datu par galaktikām iegūšana ar apakšlogu metodi**

Pirmajā posmā tiek izveidota attēlu datu bāze. Attēli ir sadalīti vairākās klasēs. Ar šo metodi ir domāta liela skaita kvadrātu izņemšana (apakšlogi) no treniņattēliem (att.3.). Dotajā metodē visi apakšlogi ir nejauša izmēra, to atrašanās vieta arī tiek izvēlēta nejauši. Katra apakšloga izmērs svārstās, sākot ar 1 pikseli līdz maksimāli iespējamam izmēram (attēla augstums vai platums) (Raphaël Marée, 2005, lpp. 2).

Pozīcija tiek izvēlēta nejauši, bet apakšlogi neiziet ārpus attēla rāmjiem. Tie var brīvi pārklāties viens ar otru. Pateicoties šai pieejai, ir iespējams samērā ātri pārklāt lielāko attēla daļu. Apakšlogu izvilkuma metodika ir universāla un der jebkuriem attēlu tipiem.



3. attēls. „Apakšlogu nedefinēšana galaktikas attēla”

Metodes testēšanas rezultāti parādīja (Raphaël Marée, 2005, lpp. 3)., ka 100 apakšlogi bieži vien ir pietiekami korektai klases noteikšanai. Pētījuma ietvaros veiktais eksperiments parāda, ka, jo vairāk apakšlogu ir nedefinēts, jo lielāka ir attēla atpazīšanas kvalitāte.

### **1.3.6. Risinājumu koki**

Daži no mašīnatpazīšanas algoritmiem pamatojas uz risinājumu koku izveidošanas (Luger, 2001, lpp. 392). Ar noteikumu tiek domāta loģiskā konstrukcija „Ja... Tad...” veidā.

Risinājumu koki ir noteikumu aplūkošana hierarhiskā, secīgā struktūrā, kurā katram objektam atbilst vienīgais mezgls, kas sniedz risinājumu (Ville, 2006, lpp. 2).

Tā kā visi objekti iepriekš tika attiecināti uz mums zināmām klasēm, tad šādu risinājumu koka izveidošanās procesū sauc par mācīšanos ar skolotāju. (supervised learning). Mācīšanas process vēl tiek dēvēts par induktīvo apmācību vai koku indukciju (tree induction) (Jones, 2008, lpp. 172).

### **1.3.7. Ansambļa metode**

Ļoti nozīmīgos jautājumos, kuros ir finanšu, medicīnas, sociālas vai citas sekas, cilvēks vēlas iegūt arī citu viedokli, pirms pieņemts lēmumu. Tādā veidā rīkojoties, viņš apsver atsevišķus viedokļus. Cilvēkam ir raksturīgi jautāt „vairākiem ekspertiem” pirms galīgā lēmuma pieņemšanas. Tomēr šī procesa priekšrocības tika izmantotas pavisam nesen automatizēto risinājumu jomā (Polikar, 2006).

Ansambļa metode ir vienlaicīga vairāku mācību modeļu izmantošana. Modeļi var tikt uzbūvēti ar jebkuras metodikas palīdzības (tādi kā neironu tīkli, risinājumu koki). Tiek uzskatīts, ka ansambļa metode ir precīzāka par jebkuru no metodēm ar vienu modeli (D. Opitz, 1999, lpp. 29).

## ***1.4. Nodaļas apkopojums***

Teorētiskā analīze parāda, ka galaktiku automātiskās klasifikācijas uzdevums var tikt atrisināts ar citās zinātnēs izmantojamo algoritmu palīdzību. Tam ir nepieciešama sistēmas trenēšana, ņemot vērā galaktiku attēlu struktūras īpatnības. Lai noteiktu, cik efektīva būs datorklasifikācija un vai ir iespējams sasniegt to precizitāti, ko sasniedz iniciatīvās zinātnisko pētnieku grupas dotā pētījuma ietvaros, tika veikts eksperiments.

## 2. Praktiskā daļa

### 2.1. Attēlu atpazīšanas algoritma testēšana

Darba teorētiskajā daļā tika minēta attēlu klasifikācijas metode „ET-DIC”. Avotu analīze parādīja, ka šī metode ir efektīva bioinformātikas uzdevumos, taču nav tikusi testēta ar astronomisko attēlu kopām.

Uzstādītajam astronomiskajam uzdevumam ir virkne īpatnību. Galaktikas veida noteikšanai vispirms ir jāorientējas pēc tās formas, kurai nav skaidru robežu. Izņemot pašu galaktiku, bildē var būt arī citi objekti (attālākas galaktikas, zvaigznes, kosmiskie putekļi).

Par praktiskā pētījuma pamatmērķi var uzskatīt Pixit atpazīšanas sistēmas, kas realizē „ET-DIC” metodi, parametru atlasu un konfigurāciju, un skaitlisku rezultātu iegūšanu par metodes darba precizitāti.

Rezultātus var uzskatīt par apmierinošiem, ja tiks sasniegta precizitāte, kas nav zemāka kā manuālajai sistematizācijai. Pozitīva rezultāta gadījumā par nepieciešamo pētījuma uzdevumu var kļūt on-line datu bāzes izveidošana automātiskās sistematizācijas rezultātu publicēšanai.

#### 2.1.1. Eksperimenta nosacījumi

Teorētiskajā daļā piedāvātās metodes efektivitātes pārbaudei tika veikta eksperimentu virkne.

Amerikāņu pētnieks Dr. Lior Shamir, kas 2009. gadā veica galaktiku tipu izpēti, piekrita automātiskās atpazīšanas eksperimentam sniegt datu bāzi ar sagatavotiem un sistematizētiem galaktiku attēliem, kura tika izmantota eksperimentā. Bāze sastāv no 560 attēliem, kas ir sadalīti trīs kategorijās: eliptisko galaktiku attēli (215 attēli), spirālveida galaktiku attēli (235 attēli) un edge-on galaktiku attēli (107 attēli). Visi uzņēmumi tika veikti projekta Sloan Digital Sky Survey ietvaros.

Attēli raksturlielumi:

- Izšķirtspēja: 120x120 pikseļi
- Attēla formāts: PNG<sup>2</sup>
- Krāsu dziļums: 24 biti (Truecolor)<sup>3</sup>

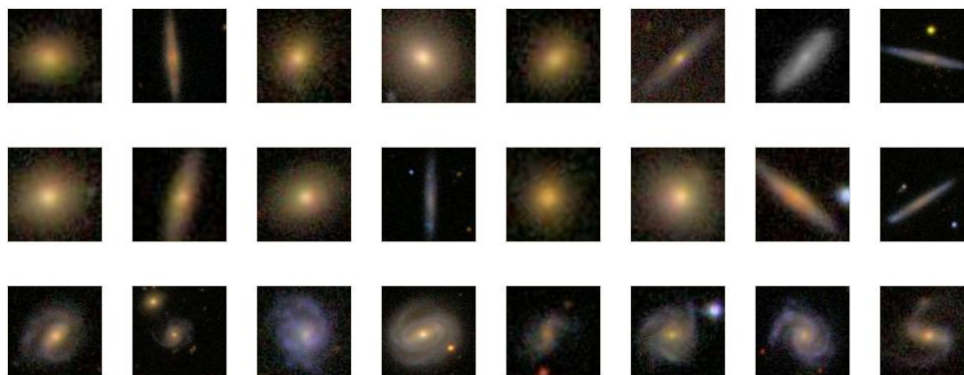
---

<sup>2</sup> Rastra grafisko attēlu uzglabāšanas formāts

<sup>3</sup> 16 777 216 iespējamās krāsas

Teleskopa, ar kuru veikti uzņēmumi, raksturlielumi:

- Debesu sfēras skata leņķis – 1.5 grādi, reflektora diametrs – 2.5m
- Observatorijas atrašanās vieta: Apache Point Observatory, New Mexico



4. attēls. „Tabulas ar sagatavotu galaktiku attēlu kopas fragments”

4. attēlā redzams, ka visi attēli ir apgriezti līdz kvadrāta formai. Tas atvieglo darbu un ļauj unificēt automātisko sistematizāciju. Eksperimentam netika izmantoti augstākas izšķirtspējas attēli, jo tas varēja palielināt sistēmas mācīšanas un testēšanas laiku.

### 2.1.2. Sistēmas darba novērtējums

Attēlu atpazīšanas sistēmu kvalitātes novērtēšanas kritērijs ir precizitāte, ar kuru notiek attēlā esošo objektu klasifikācija.

Dotajā pētījumā izmantotā klasifikācijas shēma ir pilnīgi identiska tai, kura tika izmantota projektā „Galaxy Zoo 1”. „Galaxy Zoo 1” autori savā publikācijā (Chris J. Lintott, 2008, lpp. 11) raksta, ka projekta klasifikācijas rezultātiem ir ap 10 procentu liela kļūda. Projektā piedalījās entuziasti, taču viņu darba kvalitāte nepiekāpjas datu kvalitātei, ko ieguva profesionāli zinātnieki. Šāds rezultāts ir sasniegts, pateicoties dažādu lietotāju daudzkārtējai viena attēla pārbaudei. Faktiski automātiskās attēlu klasifikācijas ansambļa metode atkārti šo praksi, testējot attēlu, izmantojot vairākus klasifikatorus.

Tādā veidā par orientējošo precizitāti ir pieņemta 10% nepareizas noteikšanas robeža, kuru sasniedzot automātiskās klasifikācijas rezultāts ir uzskatāms par neapmierinošu. Jebkura prognozi veidojoša programma var sniegt kļūdainu rezultātu. Faktiski šādu kļūdu atkārti biežums arī ir programmas darba kvalitātes rādītājs.

### 2.1.3. Atpazīšanas sistēmas uzstādījumu parametru noteikšana

#### **Attēlu segmentācija**

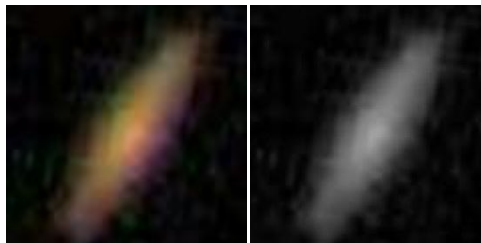
Segmentācija ir attēla sadalīšana vairākos reģionos. Tā tiek izmantota attēla vienkāršošanai, kas savukārt ļauj vienkāršot tā analizēšanu (Linda G. Shapiro, 2001).

Šīs operācijas rezultāts ir segmentu kopa, kas aptver visu attēlu vai dažu objektu kontūras (edge detection). Visiem pikseliem segmentā ir kopīgi raksturlielumi (krāsa, intensitāte, tekstūra).

#### **Etalona testēšana**

Metožu efektivitātes salīdzināšanai tiek veikti divi etalona testi. Standarta uzstādījumi (izņemot krāsas kodēšanu) nemainījās.

Konsultācijas ar metodes izstrādātāju Ljēžas Universitātes darbinieku Doktoru Marī (4. pielikums) ļāva noskaidrot, ka attēla kodēšana pelēkajā paletē var ievērojami samazināt izmantojamās operatīvās atmiņas apjomu un ievērojami paātrināt eksperimentu. Šādu uzstādījumu veidu var izmantot, ja krāsas nenes nekādu svarīgu informāciju pie objekta klases noteikšanas. Turpmākā testēšana tika veikta divos režīmos – krāsainajā un pelēkajā.



5. attēls. „Testēšana krāsu un pēlektōņu režīmos”

Pirmajā eksperimentā apakšlogi tika saglabāti kā krāsaini attēli (krāsas kodējums — HSB<sup>4</sup>). Otrā eksperimenta laikā apakšlogi tika saglabāti kā pēlektōņu attēli (5. att.). Testēšanas rezultāti parādīja, ka apakšlogu krāsainais kodējums ļauj iegūt minimālu klasifikācijas kvalitātes uzlabošanu. Tomēr tās izmantošana 10 attēlu klasifikācijai aizņem divas reizes ilgāku laiku (30 sekundes) salīdzinājumā ar pēlektōņu. Pamatojoties uz to, tika izlemts turpmākos eksperimentus veikt abos

---

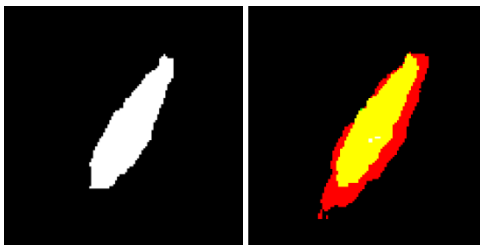
<sup>4</sup>HSB – hue, saturation, brightness (nokrāsa, piesātinājums, spilgtums). HSB modelī jebkura krāsa veidojas no spektrālās krāsas, pievienojot noteiktu procentu baltas un melnas krāsas

krāsu kodēšanas režīmos (HSB un pēlektoņu). Abi režīmi pārsniedza pieļaujamo kļūdu sliekšni, līdz ar to ir nepieciešams pārbaudīt citus programmnodrošinājuma uzstādījumus.

### ***Spilgtuma līmenis (Threshold)***

Spilgtuma līmenis ir viena no vienkāršākajiem un efektīvākajiem attēla segmentācijas metodēm. Pēlektoņu attēls tiek transformēts binārajās krāsās (melna un balta).

Apstrādājot attēlu ar krāsu sliekšni, atsevišķi pikseļi tiek atzīmēti kā „objekta” pikseļi, ja to krāsa pārsniedz kādu rādītāju (no 0 līdz 255). Pārējie kļūst par „fona” pikseļiem. Objekta pikseļi tiek atzīmēti kā „1”, bet fona pikseļi tiek atzīmēti kā „0”. Beigu etapā binārais attēls tiek izveidots no iepriekš izveidotās vieninieku un nulļu matricas. 6. attēlā pa kreisi: melnbalts spilgtuma sliekšnis, pa labi – krāsains.



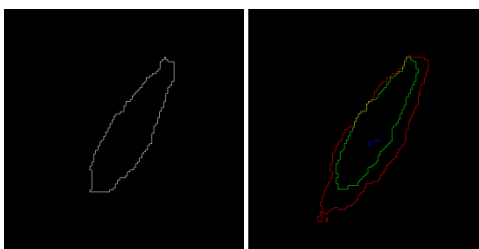
6. attēls. „Testēšana B/W un krāsu režīmos”

Testu laikā abi režīmi parādīja vienādu klasifikācijas kvalitāti. B/W (melbaltais) režīms atkal izrādījās ātrāks (20 pret 39 sekundēm krāsu režīmā). Abi rezultāti pārsniedz maksimāli pieļaujamo kļūdu koeficientu.

### ***Konvolūcijas matrica***

Konvolūcijas matrica (convolution) ir vienas matricas pārvēršana ar citas matricas, kas saucas *kodols* („kernel”), palīdzību. Attēlu apstrādē kā sākuma dati ir RGB kanālu matricas taisnstūra koordinātēs.

Uz konvolūcijas matricas balstīta filtra lietošana ļauj atdalīt objektu kontūras attēlā. Filtrs efektīvāk darbojas kopā ar krāsu sliekšņa filtru, kas ir lietots attiecībā krāsainam vai grayscale attēlam.



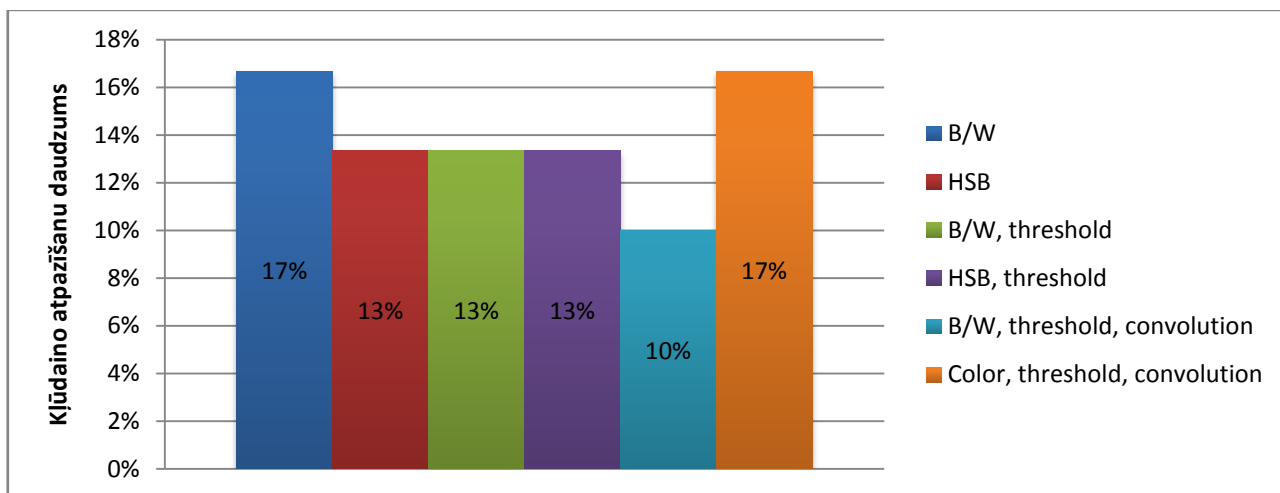
7. attēls. „Konvolūcijas”

Krāsu režīms uzrādīja lielu kļūdu daudzumu. Kontūra tika izņemta nepareizi, kas traucēja pareizas galaktikas tipa noteikšanai.

Izmantojot „grayscale” režīmu, šī problēma netika novērta. Galaktiku kontūra tika izņemta pareizi. Kļūdu koeficients sastādīja 10%, kas bija atbilstoši uzstādītajām prasībām.

### **Optimālu uzstādījumu izvēle**

Pētījuma praktiskās daļas procesā tika testētas visas uzskaitītās metodes un pēc tam salīdzinātas savā starpā.



8. attēls. „Kļūdaino atpazīšanu daudzums”

Pēc veiktajiem pētījumiem tika novērtēta kļūda, ar kuru darbojās sistēma pie dažādiem uzstādījumiem. Pētījumi parādīja, ka attēla segmentācija pozitīvi ietekmēja veicamās klasifikācijas kvalitāti. Grayscale izmantošana apakšlogu kodēšanā rekomendēja datora izmantojamus resursus. Izmantojot optimālos uzstādījumus (grayscale, threshold, convolution), sistēma klasificēja attēlus ar precizitāti, kas bija vienāda ar cilvēka veikto klasifikāciju. Visi eksperimenta rezultāti ir norādīti 3. pielikumā



### **Testēšana izmantojot izskaitļošanas klasteru**

Liela mēroga testēšanas veikšanai bija nepieciešama liela skaitļošanas jauda. Neņemot vērā, ka Pixit programmnodrošinājums var darboties uz parasta mājas datora, precīzai sistēmas pārbaudei bija nepieciešama vesela eksperimentu sērija un līdz ar to arī ilgāks laiks apstrādei.

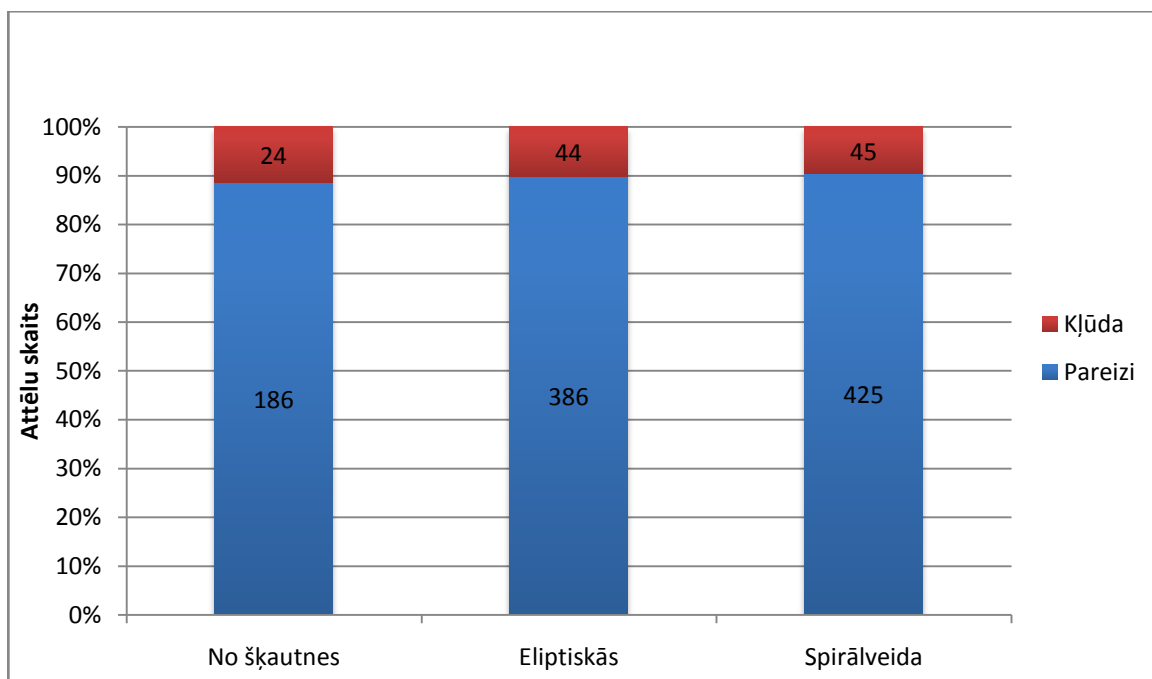
Dotā pētījuma autors sazinājās ar beļģu institūtu University of Liège un nosūtīja viņiem Latvijas testa rezultātus.

Veikto pētījumu rezultāti ieinteresēja GIGA Bioinformatics grupas dalībniekus, un viņi izdalīja 3 stundas darbam ar viņu rīcībā esošu klasteru (liela datoru grupa, kas veic paralēlas skaitļošanas) nepieciešamajiem pētījumiem.

Sistēmas apmācībai bija izmantoti 80% no attēlu bāzes un 20% tika izmantoti testēšanai. Statistiskās kļūdas samazināšanai katrs tests tika veikts 10 reizes.

Kopā ar „ET-DIC” algoritmu (dotā pētījuma pamatalgoritms) vēl tika testēta arī jaunā ET-BOF metode, kas vēl ir testēšanas stadijā.

### **ET-DIC metode**



9. attēls. „ET-DIC testēšana”

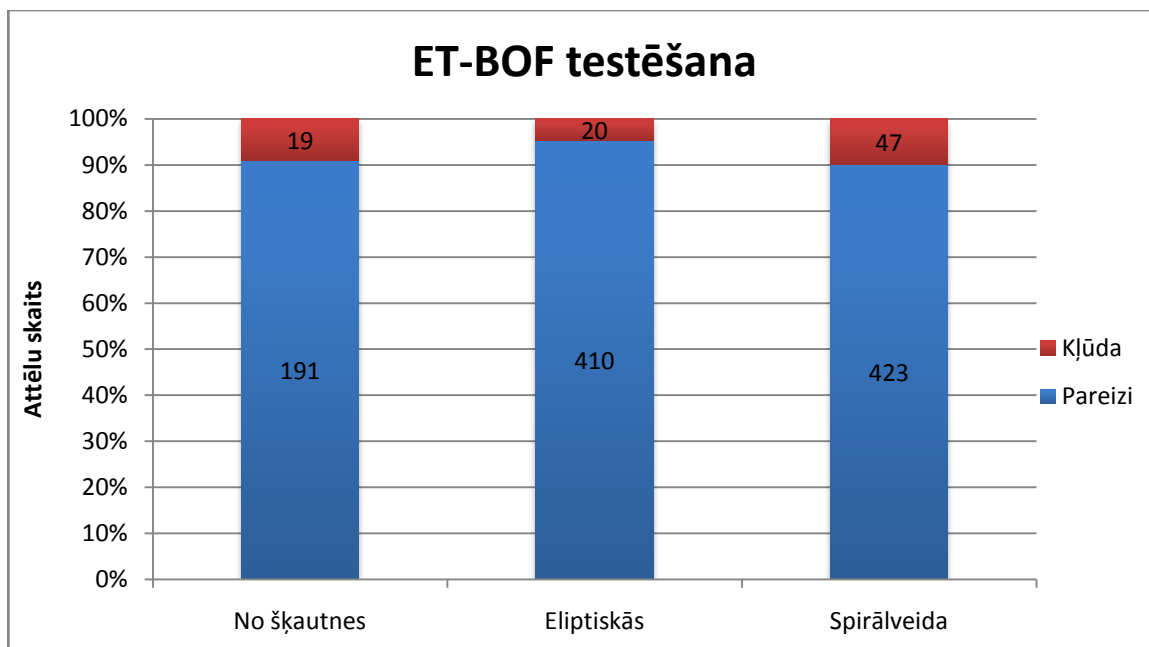
Apakšlogu izmērs bija mainījies robežās starp 25% un 50%.Pēc testēšanas rezultātiem tika izskaitļots kļūdu koeficients - 10.19 +/- 2.06%.Iegūtie dati atbilst iepriekš Latvijā veiktās testēšanas

datiem. Kļūdu koeficients nepārsniedza pieļaujamo robežu, kas pierāda testējamās metodes pilnvērtību.

### **ET-BOF metode**

ET-BOF ir jauna attēlu klasificēšanas metode, ko testē zinātnieki no GIGA Bioinformatics. Tā balstās uz iepriekš aprakstīto ET-DIC metodi, taču tai ir daudz ievērojamu uzlabojumu.

Dotajā momentā ET-BOF metode atrodas aktīvas pārbaudes stadijā.



10. attēls. „ET-BOF testēšana”

Labākie rezultāti tika iegūti, izmantojot apakšlogus ar izmēru 10-50% no attēla izmēra. Dotās metodes kļūdas koeficients – 7.75 +/- 2.88 %.

#### **2.1.4. Testēšanas rezultāti**

Rezultāti, kas iegūti veicot rūpīgāku testēšanu, nedaudz atšķiras no tiem, kas tika iegūti pirmreizējās testēšanas rezultātā. Tomēr kļūda atrodas statistiskās kļūdas robežās. Statistiskā kļūda tika samazināta veicot testu sēriju uz klastera. Pētījumu gaitā izmantotā ET-DIC metode uzrādīja prognozēšanas precizitāti līdz 90%, kas atbilst cilvēka veiktajai klasifikācijai.

Testētā ET-BOF, kas ir ET-DIC atjaunotā versija, klasifikācijā izrādījās precīzāks (97% gadījumu nostrādāja pareizi). Dotā metode galaktikas klasificē precīzāk par cilvēku.

## **2.2. Rezultātu publikācija on-line datubāzē**

Praktiskās daļas ietvaros tika uzbūvēts modelis, kas ļauj klasificēt dažus galaktiku tipus. Vairums teoriju izmanto datus par galaktiku morfoloģisko struktūru. To pierādīšanai ir nepieciešams liels datu apjoms.

Darba praktiskās daļas ietvaros, pamatojoties uz izvēlēto klasifikācijas modeli, tika izveidota mājas lapa ar publiski pieejamu galaktiku datu bāzi (1. pielikums), kas satur informāciju par to morfoloģisko struktūru <http://www.rkg.lv/projectdebess/>. Informācija par datu bāzi ir publicēta zinātnisko apvienību, kas nodarbojas ar šo problēmu, forumos, lai tādā veidā dati būtu pieejami pētniekiem. Pētnieciskā darba beigšanas brīdī datu bāzē ir vairāk par 500 ar E-DIC algoritmu klasificētu attēlu, un to skaits turpina pieaugt.

Projektā ir realizēta vairums funkciju, kas raksturīgas „crowdsourcing”<sup>5</sup> projektiem. Ir funkcija „paziņot par kļūdu”, pēc kuras izsaukšanas lietotājs var norādīt pareizu objekta klasi. Pēc kritiska balsu daudzuma sasniegšanas (5 lietotāju balsis) objekts automātiski nomaina savu klasi (ir paredzēta aizsardzība pret tīklu robotiem). Tas ļauj samazināt kļūdu daudzumu datu bāzē, kuru tiks izmantos arī citos pētījumos.

Attīstoties attēlu bāzei, tiks veikta jaunu modeļu mācīšana. Sistēma uzlabos klasifikācijas kvalitāti uz treniņiem paredzēto attēlu daudzuma palielināšanās rēķina, un, pateicoties lietotāju palīdzībai, samazināsies viltus objektu daudzums treniņiem paredzētajā bāzē.

---

<sup>5</sup>Darbus dara brīvprātīgie profesionāļi par to nesaņemot finansiālu atlīdzību.

## Rezultāti

- Notikušas intervijas ar speciālistiem datorredzes jomā. Tas jāva noteikt grūtības, kuras varēja rasties izvīrzītā uzdevuma risināšanas gaitā.
- Eksperimenta veikšanai tika izvēlēta Pixit programmu pakete, kas realizē ET-DIC metodi un ir adaptēta pētījuma uzdevumam.
- No 2009. gada novembra līdz 2010. gada janvārim veikta uzstādītās programmu paketes konfigurēšana, trenēšana un testēšana.
- Vislabākos automātiskos atpazīšanas rezultātus (Rīgas eksperimenta ietvaros) izdevās iegūt, izmantojot attēla segmentāciju ar krāsu sliekšņa un konvolūcijas matricas metodēm, izmantojot tās uz galaktiku attēlu etalonu kopas. Rezultāti tika salīdzināti ar atvērtā starptautiskā projekta Galaxy Zoo datiem. Eksperiments pierādīja, ka automātiskā galaktiku atpazīšana nav sliktāka par manuālo.
- Veiktā eksperimenta rezultāti jāva iesniegt iesniegumu University of Liège (Beļģija) un iegūt, stundu sistēmas darba grantu tīkla klasterī (ekvivalents PC darba 300 stundām), pateicoties kam, tika iegūti plaši dati par automātiskās sistematizācijas precizitāti. Beigās tika sasniegti 93% pareizas atpazīšanas precizitātes.
- Beļģijā testētais jaunais eksperimentālais attēlu klasifikācijas algoritms deva 3% precizitātes pieauguma salīdzinājumā ar iepriekšējo algoritma realizāciju.
- Iegūtie dati apstiprināja Rīgā šī pētījuma autora Rīga veiktā eksperimenta rezultātu. Precizēti sistēmas uzstādījumu, koku dziļuma un apakšlogu parametri, kas sniedz labāku rezultātu. Eksperiments ar lielu datu kopu pierādīja, ka automātiskā sistematizācija notiek ar vismaz 90% precizitāti.
- Izveidots Web-resurss, kas orientēts uz automātiskās un manuālās klasifikācijas kļūdu atrāšanas projekta uzturēšanu un attīstību. Resursā izlikta kopa no 1000 galaktiku attēliem, kas ir klasificēti pa tiem.

## Secinājumi

- Astronomijas zinātnisko problēmu risinājums, izmantojot datoru metodes, ir aktuāls mūsdienu uzdevums. Tā aktualitāte ir palielinājusies īpaši pēdējos gados sakarā ar radiosignālu no kosmosa, orbitālā teleskopa Hubble neapstrādāto datu apjoma palielināšanos un tādu projektu kā Sloan Digital Sky Survey aktivitāti. Līdzīgi projekti ir liela daudzuma Visuma attālāko nostūru ciparu uzņēmumu piegādātāji.
- Viena no svarīgākajām fundamentālās astronomijas problēmām ir galaktiku klasifikācija. Šobrīd šis jautājums tiek risināts manuāli, taču var tikt atrisināts ar lietisķās informātikas līdzekļiem.
- Bioloģijā, optikā, inženieru un ofisu uzdevumos izmantojamās attēlu atpazīšanas datormetodes var tikt veiksmīgi adaptētas sistematizācijas problēmas atrisināšanai astronomijā.
- Ansambļu algoritms un apakšlogu algoritms ir visperspektīvākie instrumenti galaktiku attēlu atpazīšanas jomā, tomēr tiem ir nepieciešama smalka apmācība, izmantojot etalona attēlu kopu.
- 10% kļūdas pastāvēšana automātiskajā klasifikācijā nav problēma dotajam uzdevumam. Fundamentālajai astronomijai ir svarīga informācija par galaktiku tipiem lielos klasteros. Iegūtā precizitāte ir pieļaujama Visuma attīstības modeļu veidošanai plašā mērogā.
- Pētījuma rezultāti parādīja, ka galaktiku klasifikācijas algoritmi ar uzstādīto uzdevumu tiek galā ne sliktāk kā cilvēks un ievērojami palielina sistematizācijas ātrumu. Rodas iespēja veikt vēl lielāka mēroga pētījumus, jo datu apstrādes spējas nav atkarīgas no projektā iesaistīto cilvēku skaita. Tādā veidā darba mērķi var uzskatīt par sasniegtu un hipotēzi – par pierādītu.
- Astronomijas problēmas, kuras atrisināt var palīdzēt informātika, neaprobežojas vienīgi ar galaktiku klasifikācijas uzdevumu. Šādu uzdevumu ir pietiekošami daudz, un vairumam joprojām nav atrasti risinājumi. Dotais pētījums nemeclē atbildi uz jautājumu, kā pareizi rīkoties ar iegūto informāciju par galaktiku tipiem. Šāda veida secinājumus ir tiesīgi izdarīt vienīgi profesionāli zinātnieki, astronomi. Tomēr atrisinātais uzdevums, lai arī liekas viegls, tomēr var ievērojami „atvieglot dzīvi” pētniekiem, kas nodarbojas ar Visuma vecuma un attīstības jautājumiem.

## **Noslēgums**

Lai arī informātika ir augstākās pakāpes lietišķā disciplīna, tās uzdevumi neaprobežojas vienīgi ar uzdevumiem „kā palielināt datu bāzes drošību” vai „kā uzģenerēt fotorealistisku ainavu filmai par citplanētiešiem”. Informātika var un arī veiksmīgi risina fundamentālo zinātņu – matemātikas, astronomijas un citu – problēmas. Tās ir atbildes uz jautājumiem, kuras nevar sniegt zinātne datu lielā apjoma un objektu kombināciju lielā skaitļa dēļ. Var teikt, ka informātika un datorzinātnes ir pilnvērtīgi dalībnieki atbilžu meklēšanā uz cilvēces pamatjautājumiem – „kas mēs esam”, „kā ir veidota pasaule mums apkārt”, „no kurienes viss ir radies un par ko tas var pārtapt”.

## **Izmantotas literatūras saraksts**

- Anil K. Jain, R. P.** (2000). Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis & Machine Intelligence, lpp. 4-37.
- Chris J. Lintott, K. S.** (2008). Galaxy Zoo : Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey.
- D. Opitz, R. M.** (1999). Popular Ensemble Methods: An Empirical Study. Journal of Artificial Intelligence Research, lpp. 169-198.
- Eberly College of Science.** (2009). Astrostatistical Challenges. Ielādēts 2010. gada 28.02 no <http://www.astrostatistics.psu.edu/astchall.html>
- Elske V.P. Smith, K. C.** (1987). Introductory Astronomy and Astrophysics. Saunders College Publishing/Harcourt Brace.
- Eremeeva A.I., Cicins F.A.** (1989). Astronomijas Vēsture. Maskava: MGU.
- Krieg, S.** (2005). Kepler's Case, „GEO” nr.11., lpp. 150-166.
- Linda G. Shapiro.** (2001), Computer Vision. Pearson: US. lpp. 315-337.
- Luger, G. F.** (2001). Artificial Intelligence: Structures and Strategies for Complex Problem Solving (4th Edition). Reading: Addison Wesley.
- Polikar, R.** (2006). Ensemble based systems in decision making. Circuits and Systems Magazine, IEEE, 6 (3), lpp. 21 - 45.
- Raphaël Marée, P. G.** (2005). Random Subwindows for Robust Image Classification. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. San Diego. lpp. 7.
- Raphaël Marée, Pierre Geurts, Louis Wehenkel.** (2006). Biological Image Classification with Random Subwindows and Extra-Trees. Bio-Image Informatics (Workshop on Multiscale Biological Imaging, Data Mining & Informatics), Santa Barbara. lpp. 2.
- Simon, H. A.** (1983). Search and Reasoning in problem solving. Artificial Intelligence, lpp.7-29.
- William J. Kaufmann, I.** (1977). Astronomy: The Structure of the Universe. New York: Macmillan.

## **Anotācija**

Pētnieciskais darbā ir aplūkotas attēla datoatpazīšanas metodes fundamentālās astronomijas zinātniskā jautājuma – galaktiku klasifikācijas kontekstā. Darbā notiek izvēlētās tēmas teorētiskie un praktiskie pētījumi, tiek veikts atpazīšanas algoritmu precizitātes skaitliskā noteikšana. Pētījuma starprezultāti bija pamats iesniegumam par pieejas saņemšanu lielam skaitļošanas tīklam vienā no beļģu skaitļošanas centriem. Tādā veidā iegūtās skaitļošanas jaudas ļāva veikt daudzpusēju galaktikas klasifikācijas metožu testēšanu. Pēc darba rezultātiem internetā ir izveidota datu bāze turpmākai projekta attīstībai. Darba rezultāti var būt interesanti pētniekiem informātikas jomā, kas nodarbojas ar objektu atpazīšanas problēmu risināšanu un pētniekiem astronomijas jomā, kam ir nepieciešami lielapjoma dati par galaktiku tipiem.

## **Abstract**

Research work overviews methods of computer image recognition, applying them to tasks of fundamental astronomy. Theoretical and practical research are held. Quantity values of research results of recognition accuracy are being measured. Intermediate results of research has given an opportunity to win a research grant to access a large computational network in one of Belgium's research facilities. Computational powers are being used for wide testing of the methods of automatic galaxy classification. Website with a database of classified galaxies is being designed to encourage project's future development. Research results may be useful to computer scientists who work in image recognition field, as well as astronomers, who need large data amounts about galaxy types.

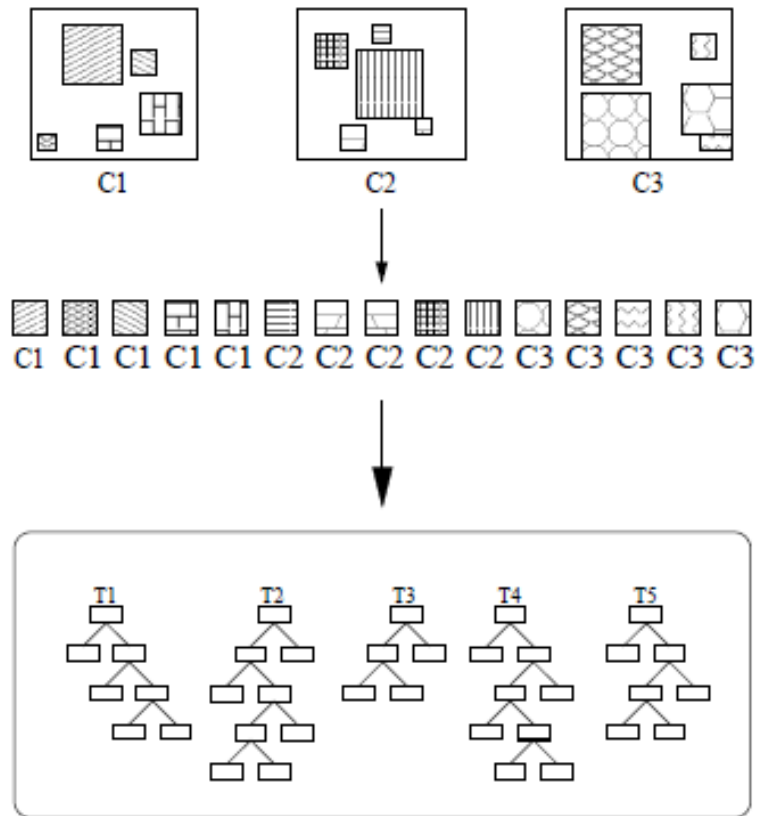


## Pielikumi

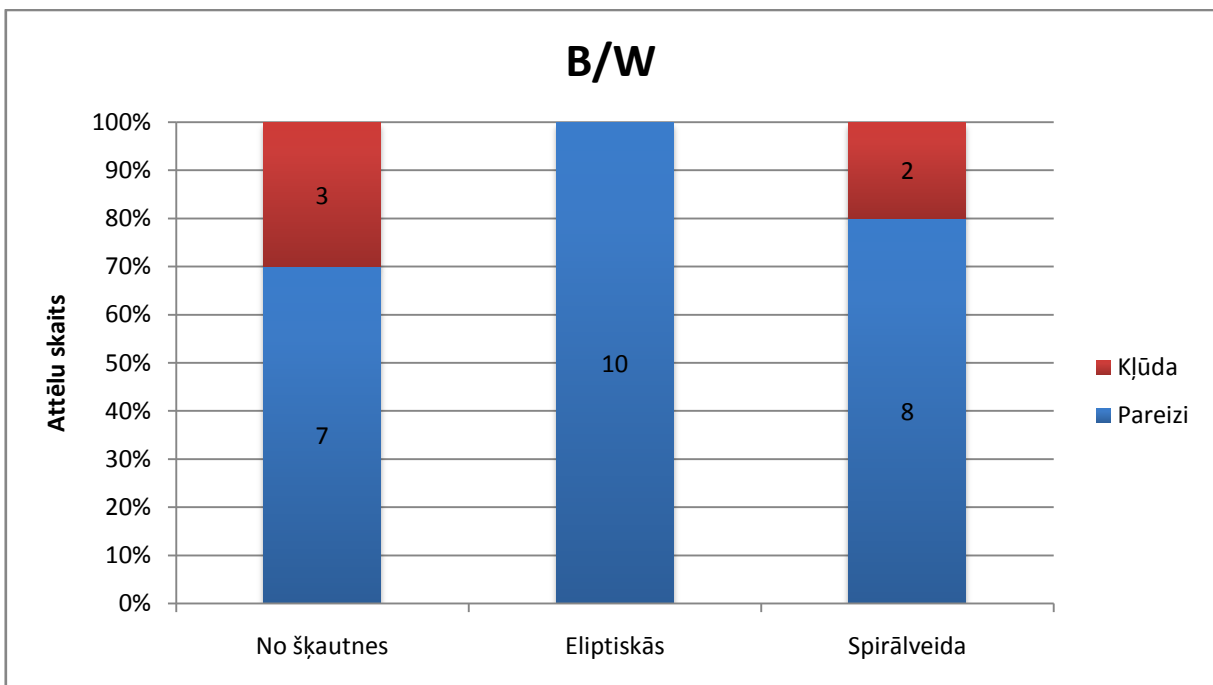
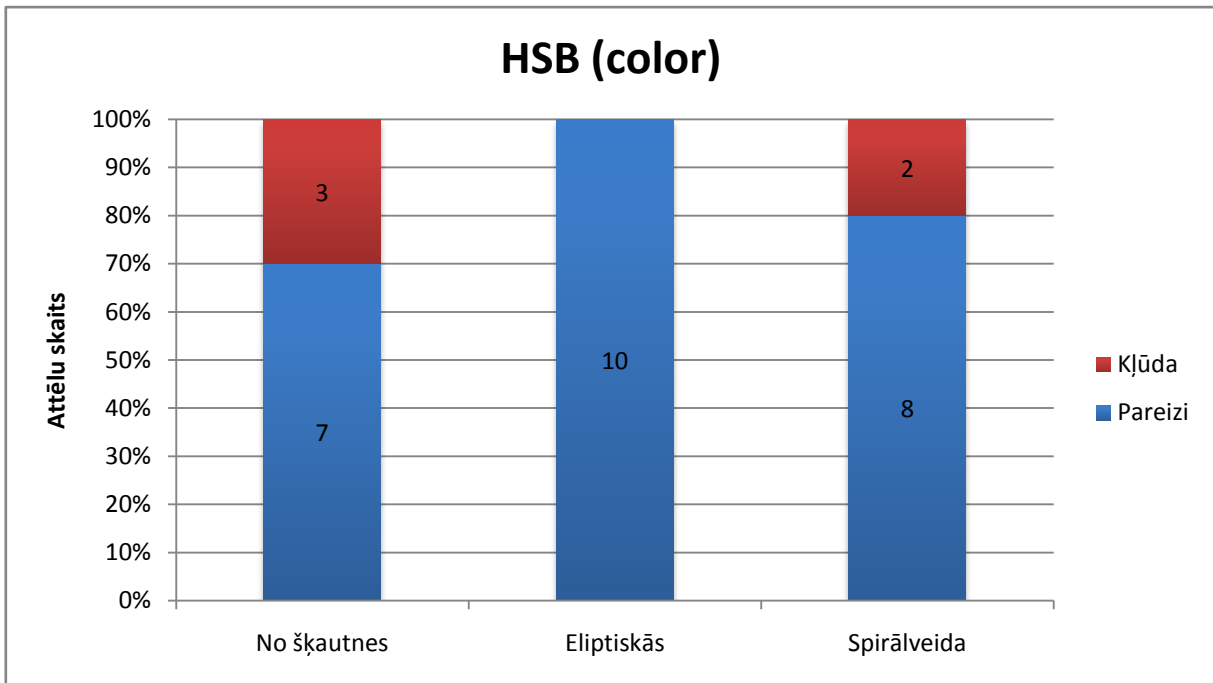
1. Pielikums. Praktiskās daļas ietvaros izveidota mājas lapa ar publiski pieejamu galaktiku datu bāzi

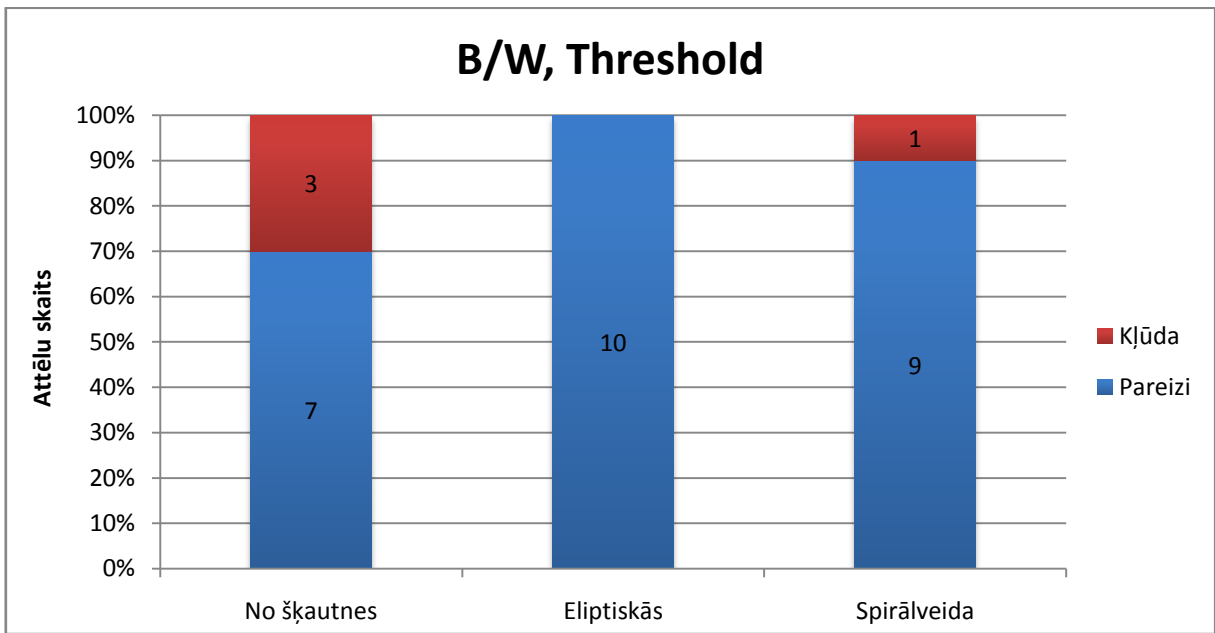
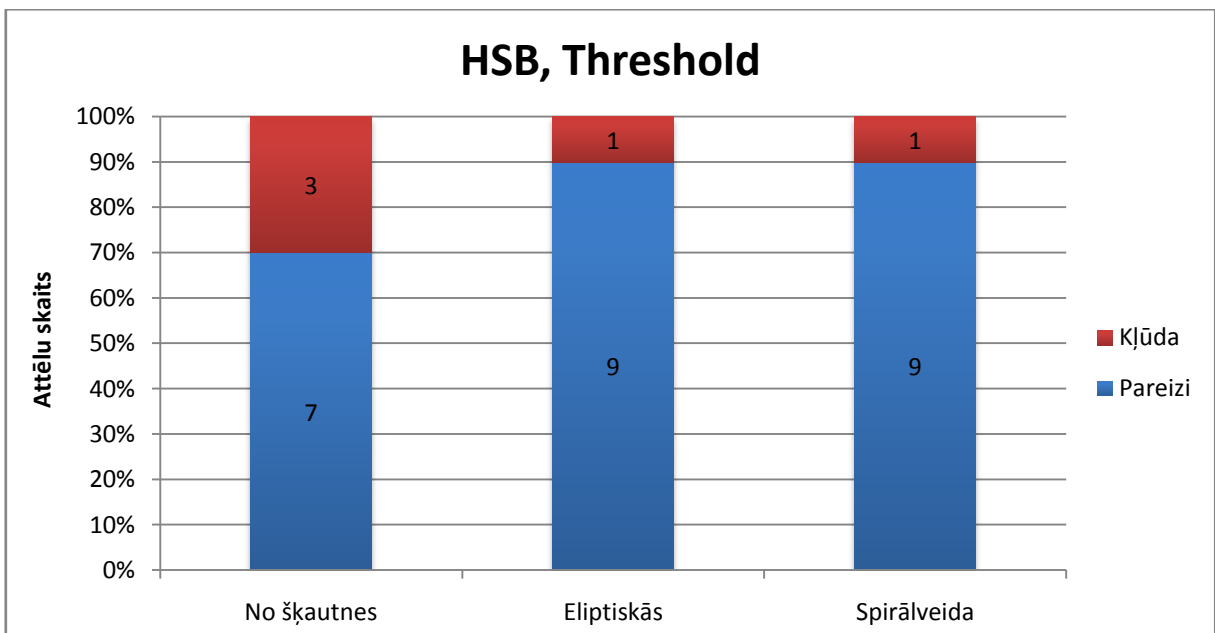


2. Pielikums. Metode ET-DIC (Raphaël Marée, 2005, lpp. 2)

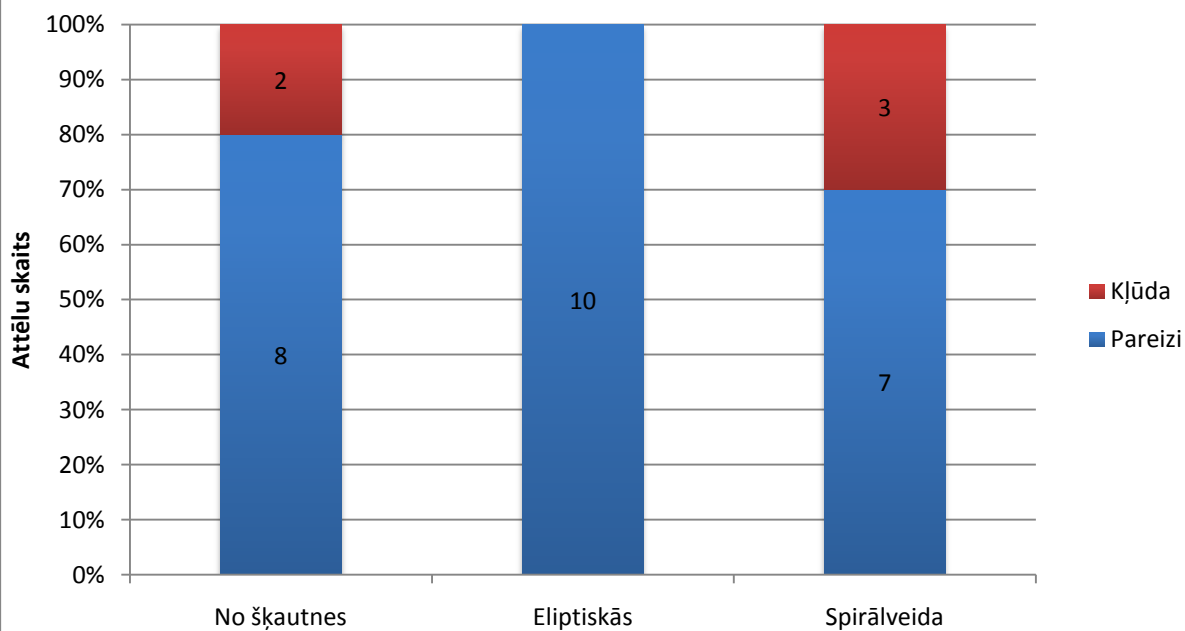


### 3. Pielikums. Testu rezultāti, izmantojot dažādu metožu kombinācijas

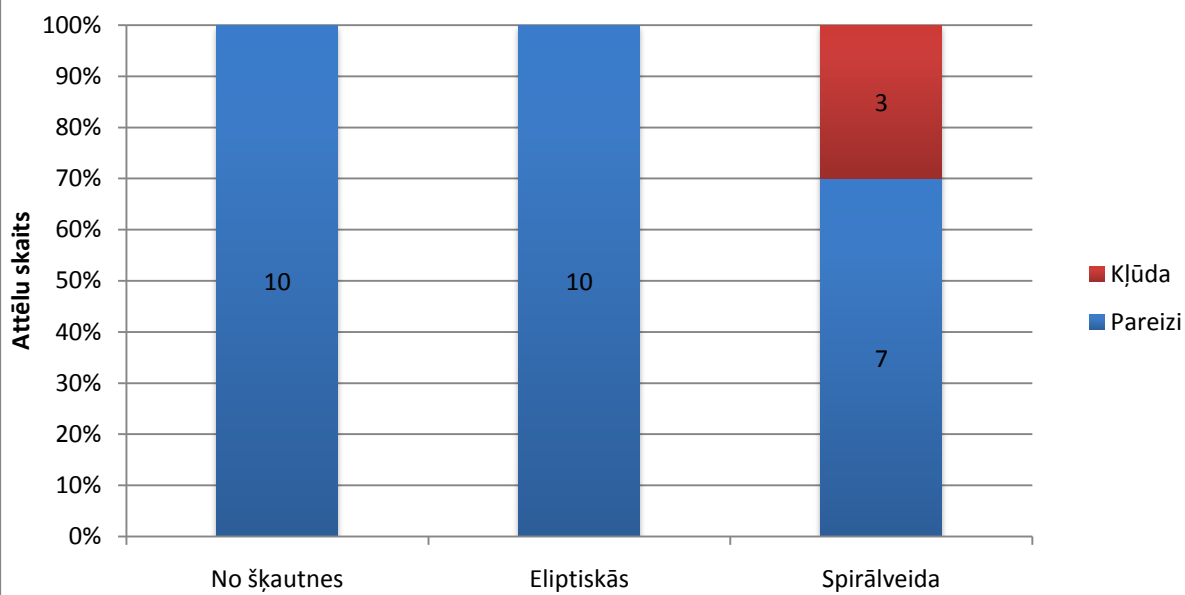




### Color, threshold, convolutuion



### B/W, threshold, convolutuion



#### 4. Pielikums. Intervija ar Dr. Raphaël Marée

##### **Raphaël Marée**

GIGA Bioinformatics Core Facility

Bioinformatics and Modeling, GIGA Research

Département d'Électricité, Électronique et Informatique

Liège, Belgique

##### **Daļa 1**

Sazonov: I'm a student of Riga Classical Gymnasium and currently doing a scientific project, which combines informatics and astronomy. The goal is to implement a system which would do automatic Hubble classification of galaxies. At this point of my research, I'm looking for a image classification method, that can handle the task. In your works, you've presented an image classification method ET-DIC, which, in some researches, was used to classify biological objects. The idea was to use biological classification software, because of the similar (to galaxies) morphology of some bio objects (e.g. cells).

Marée: We can provide help and a free evaluation version in Java of the method ET-DIC (but I would probably not be as good as ET-BOF-frequency). We also plan to deliver an evaluation version of ET-BOF-frequency as soon as it is published but probably not before mid-2010.

Our original method ET-DIC is published in BMC Cell Biology 2007. The method that obtained the best results on the poster (ET-BOF-FREQUENCY) is not yet published (I'm preparing a very large study).

Sazonov: ET-BOF-frequency shows really fascinating results, but my project must be done before mid-2010. I would like to ask you to provide an evaluation version of ET-DIC.

##### **Daļa 2**

Sazonov: Have you tried to use any image advanced segmentation techniques combined with your method? Could it improve the accuracy?

Marée: I combined the method with convolution filters to detect lines, contours, etc. Sometimes it improves results but sometimes not.

Sazonov: Is it rational to use "black & white" option of Pixit. From my point of view, it will improve classification accuracy. Probably, this will reduce the amount of unneeded information (in this case, colors of galaxies), during the model building.

Marée: Did you mean "black & white" \*color coding\*? If color is not useful to classify, then it's a good idea to use black&white (grey levels) to reduce memory requirements.

### **Data 3**

Sazonov: Pixit showed really nice error rate. Was it previously used in any astronomical tasks?

Marée: Not at all. It was originally designed with a "generic" perspective, it was evaluated on very diverse data sets but I did not have access to an astronomical dataset.