

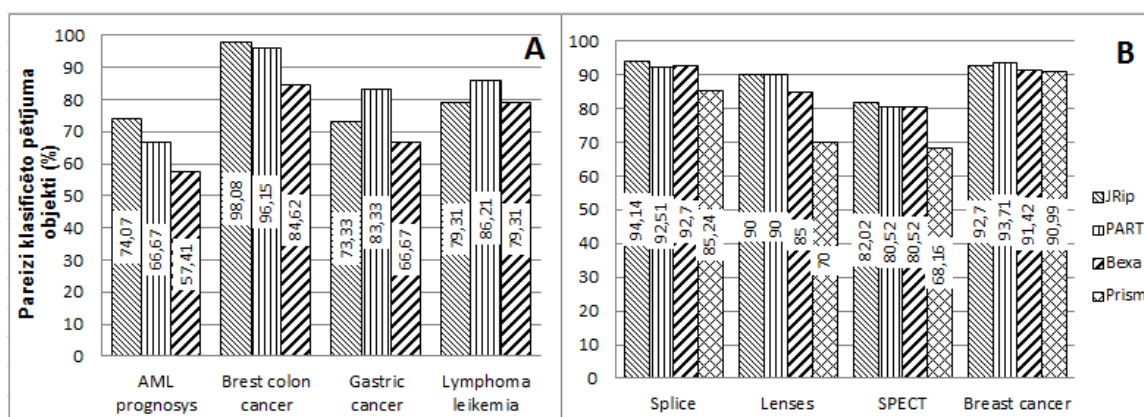
## INDUKTĪVĀS APMĀCĪBAS METOŽU IZMANTOŠANA DATU KLASIFIKĀCIJĀ

Mūsdienās ir pieejami daudzi algoritmi datu klasifikācijai, tomēr nevienu no tiem nevar uzskatīt par pilnīgu. Līdz ar to ir nepieciešams apbēt jau esošos datu klasifikācijas algoritmus un izstrādāt jaunus.

Darba mērķis - izpētīt inductīvās klasifikācijas algoritma BEXA pielietojšanas iespējas reālu bioinformātikas datu klasifikācijā.

Darba gaitā tika veikta literatūras analīzē, pētīti BEXA algoritma teorētiskie pamati, izstrādāta specializēta programmatūra datu klasifikācijas veikšanai, izmantojot BEXA algoritmu. Kā arī tika izmantota atvērtā pirmkoda programmatūra WEKA, salīdzinošo algoritmu darbības analīzei. Datu klasifikācijas algoritmi tika pārbaudīti uz vairākām reālām bioinformātikas datu kopām (nosaukumus skatīt 1.attēlā). Katrā eksperimentā tika veikta desmitkārtīga šķērsvalidācija.

Izpētot inductīvās klasifikācijas algoritma BEXA darbības principus reālās bioinformātikas datu kopās redzams, ka algoritmam ir tendence klasificēt datus ar skaitliskām vērtībām ar salīdzinoši zemāku precizitāti (vidēji par 10-15%) kā mūsdienās labāk zināmajiem algoritmiem JRip un PART (Att. A). Savukārt, datus ar nominālām vērtībām algoritms BEXA klasificē līdzvērtīgi JRip un PART algoritmiem, turklāt, būtiski labāk kā algoritms PRISM (Att. B).



**1.Att. Pareizi klasificēto pētījuma objektu skaits (%) skaitliskām (A) un nominālām vērtībām (B).**

No rezultātiem iespējams secināt:

1. Pašreizējie rezultāti liecina, ka BEXA algoritms ir piemērotāks datu klasifikācijai ar nominālām vērtībām.
2. Balstoties uz iegūtiem rezultātiem plānots izpētīt un izstrādāt algoritma BEXA izplūdušās versijas (*FuzzyBEXA* un *FuzzyBEXA 2*), kuras potenciāli varētu būt piemērotākas bioinformātikas datu analīzei.