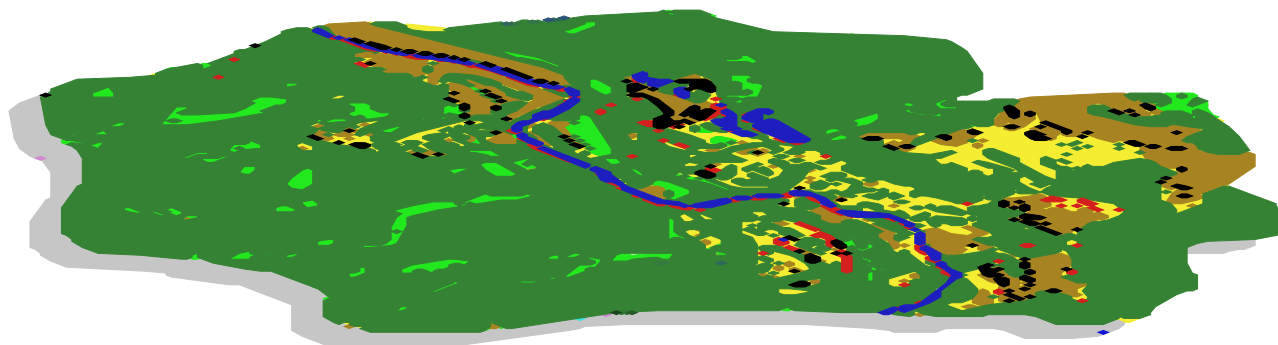


Evolúcia bázy pravidiel fuzzy klasifikátora určeného na klasifikáciu multispektrálnych satelitných obrazových dát

PREZENTÁCIA ZADANIA Z PREDMETU FUZZY ROZHODOVANIE



Michal Puheim

Obsah prezentácie

- Úvod do témy
- Evolúcia pravidiel fuzzy klasifikátora
 - Popis pravidlového systému
 - Inferenčný mechanizmus fuzzy klasifikátora
 - Iteratívne učenie pravidiel
 - Algoritmus tvorby nových pravidiel
- Multispektrálne satelitné obrazové dáta
- Implementácia systému
 - Parametre inferenčného systému fuzzy klasifikátora
 - Úpravy pôvodného návrhu plynúce z potreby zlepšenia klasifikácie
- Experimenty
- Záver

Úvod do témy

Referenčný článok:

STAVRAKOUDIS, D. G., GALIDAKI G. N., GITAS I. Z., THEOCHARIS J. B.:
***Enhancing the Interpretability of Genetic Fuzzy Classifiers in Land Cover
Classification from Hyperspectral Satellite Imagery.***

WCCI 2010 IEEE World Congress on Computational Intelligence, Barcelona, Spain. 2010.

- Iteračné učenie pravidiel pomocou GA
- Zjednodušenie pravidiel
- Optimalizácia pravidiel pomocou genetického tuningu

Ciele zadania:

- Klasifikácia pomocou fuzzy klasifikátora
- Evolúcia bázy pravidiel pomocou genetického algoritmu
- Aplikácia na multispektrálnych satelitných dátach

Evolúcia pravidiel fuzzy klasifikátora

- Klasifikátor – algoritmus priradenia objektov do tried podľa ich popisu
- Fuzzy klasifikátor – proces klasifikácie sa opiera o fuzzy logiku
- Učenie fuzzy klasifikátora – vytvorenie bázy pravidiel klasifikácie
 - Pomocou znalostí experta
 - Pomocou genetického algoritmu

Popis pravidlového systému

Fuzzy klasifikačný systém → množina fuzzy produkčných pravidiel

Pravidlo → mapovanie z príznakového priestoru do priestoru tried

$$x = \{x_1, \dots, x_N\} \in F^N \longrightarrow C = \{C_1, \dots, C_M\}$$

Tvar pravidla:

$$R^k: \text{IF } x^1 \text{ is } A_1^k \text{ and } \dots x^N \text{ is } A_N^k \text{ THEN } y \text{ is } C^k \text{ with } r^k$$

Istota pravidla:

$$r^k = \sum_{p|c^p=C^k} \mu^k(x^p) / \sum_{p=1}^Q \mu^k(x^p) *$$

* pre danú množinu označených vstupných vzoriek $E = \{(x_p, c_p), p = 1, \dots, Q\}$

Inferenčný mechanizmus FK

Vstupná premenná → Lingvistický term

Fuzzy množiny využívajú trojuholníkové funkcie príslušnosti

Pravidlo nemusí zahŕňať všetky premenné

Miera zhody pravidla:

$$\mu^k(x^p) = \mu_1^{L_q}(x_1^p) \wedge \dots \wedge \mu_N^{L_q}(x_N^p) \quad *$$

T-norma \wedge implementovaná pomocou operátora minima

Určenie triedy vzorky:

$$C^k = \arg \max_{j=1, \dots, M} \sum_{R_k | C^k = j} \mu^k(x^p) \cdot r^k$$

* $\mu_i^{L_q}(x_i^p)$ je stupeň príslušnosti lingvistického termu v predpokladovej časti pravidla k jeho fuzzy množine

Iteratívne učenie pravidiel

Použitie algoritmu AdaBoost

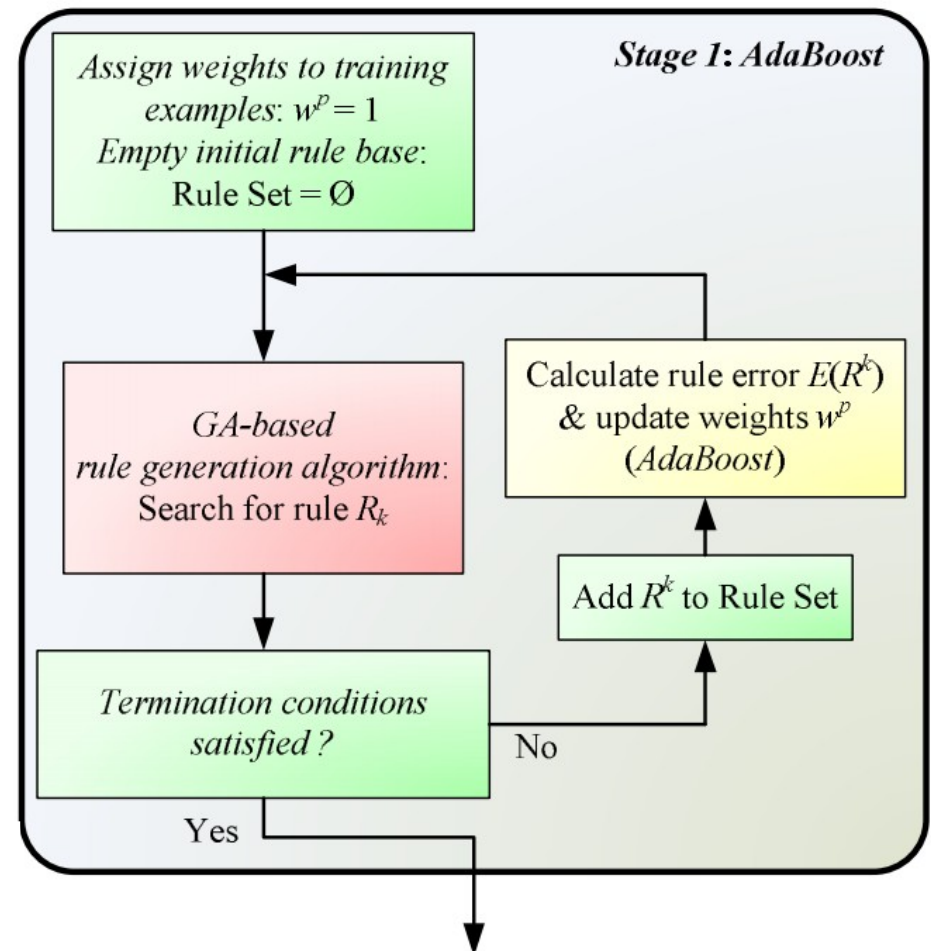
Zmena váh trénovacích vzoriek:

$$w^p \leftarrow \begin{cases} w^p, & \text{ak } c^p = C^k \\ w^p \cdot \beta^k(x^p), & \text{ak } c^p \neq C^k \end{cases}$$

$$C^k = \arg \max_{j=1, \dots, M} \sum_{p|c^p=j} \mu^k(x^p)$$

$$\beta^k(x^p) = (E(R^k) - 1 / E(R^k))^{\mu^k(x^p)}$$

$$E(R^k) = \sum_{p|c^p \neq C^k} w^p \mu^k(x^p) / \sum_{p=1}^Q w^p \mu^k(x^p)$$



Algoritmus tvorby pravidiel (1)

Genetický algoritmus – jedinec v populácii reprezentuje pravidlo
Chromozóm jedinca – rozdelený na časť premenných a hodnôt



- Elitársky model
- Pravdepodobnosť reprodukcie určená lineárne podľa „fitness”
- Výber jedincov realizovaný pomocou rulety
- Jednobodové kríženie v oboch častiach chromozómu
- Inicializácia:
 - časť premenných náhodne
 - č. hodnôt podľa trén. množiny (berie v úvahu váhy vzoriek)

Algoritmus tvorby pravidiel (2)

Funkcia vhodnosti – koncept konzistencie a pokrytia trénovacích vzoriek

Konzistencia pravidla – vysoký poč. pozitívnych vzoriek oproti negatívnym

$$\text{Cons} = \begin{cases} 0, & \text{ak } n^+ < n^- \\ 1 - n^-/n^+, & \text{inak,} \end{cases}$$

$$n^- = \sum_{p|c^p \neq C^k} w^p \mu^k(x^p)$$

$$n^+ = \sum_{p|c^p = C^k} w^p \mu^k(x^p)$$

Pokrytie – poč. vzoriek pokrytých pravidlom / poč. vzoriek danej triedy

$$\text{Cov} = n^+ / \sum_{p|c^p = C^k} w^p$$

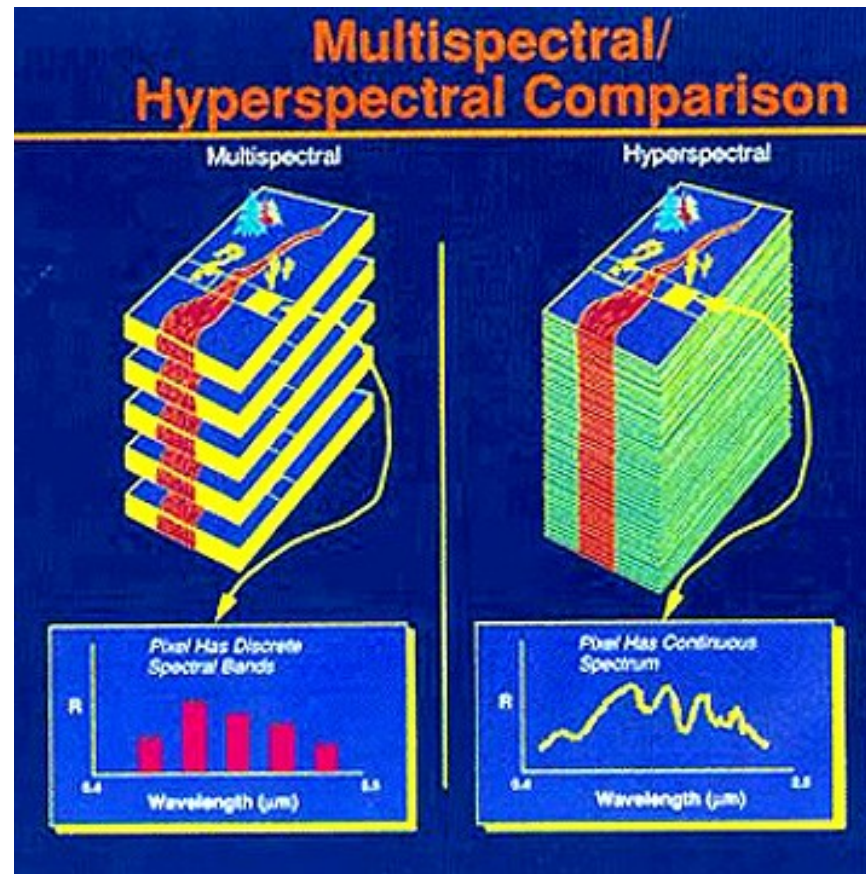
Vzťah pre výpočet funkcie vhodnosti („fitness“):

$$f_1 = \text{Cons} \cdot \text{Cov}$$

Multispektrálne obrazové dáta

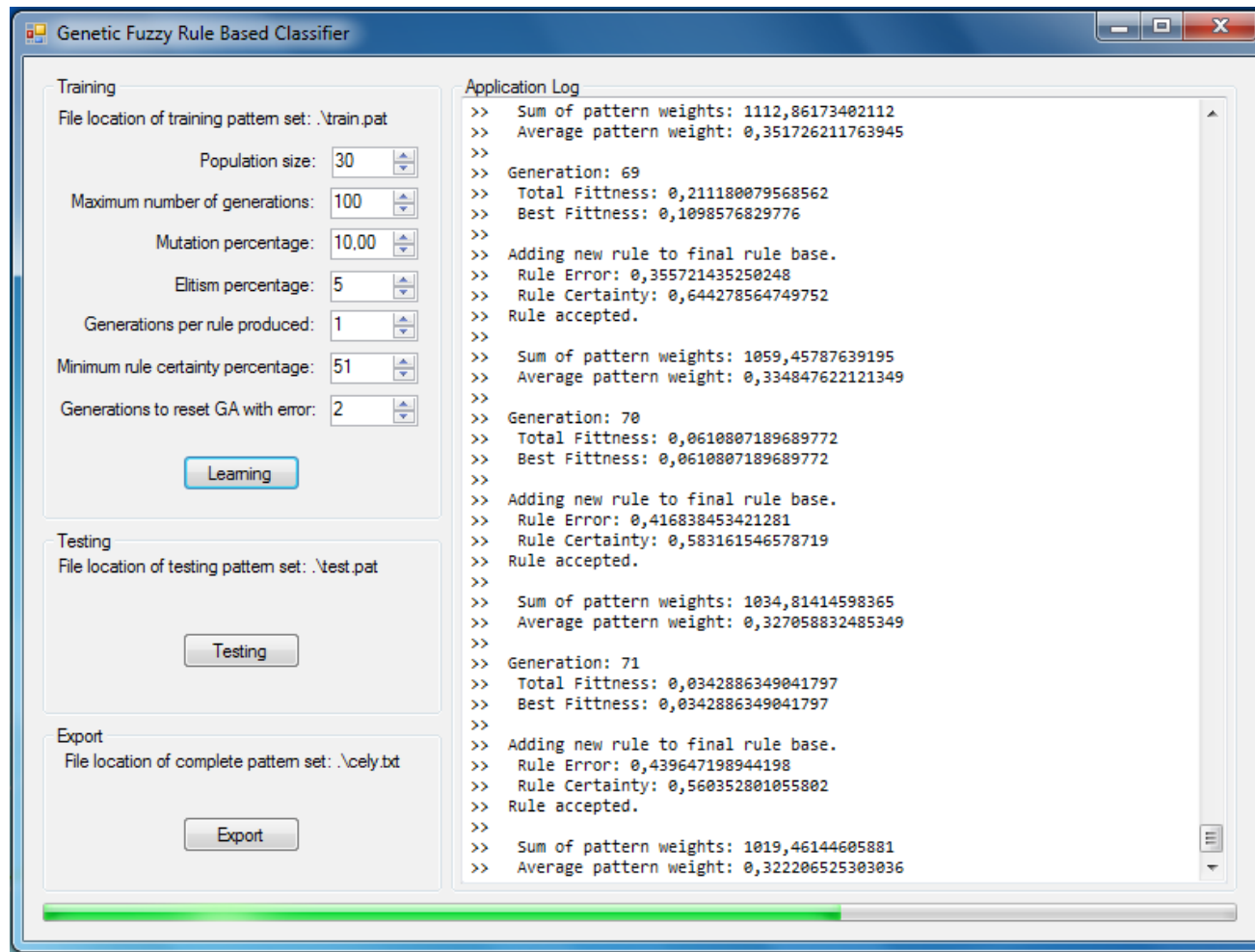
Multispektrálny – pokrýva viacero nespojitých častí el.mag. spektra

Hyperspektrálny – pokrýva súvislú časť el.mag. spektra



Implementácia systému

V prostredí MS Visual C# s grafickým užívateľským rozhraním.

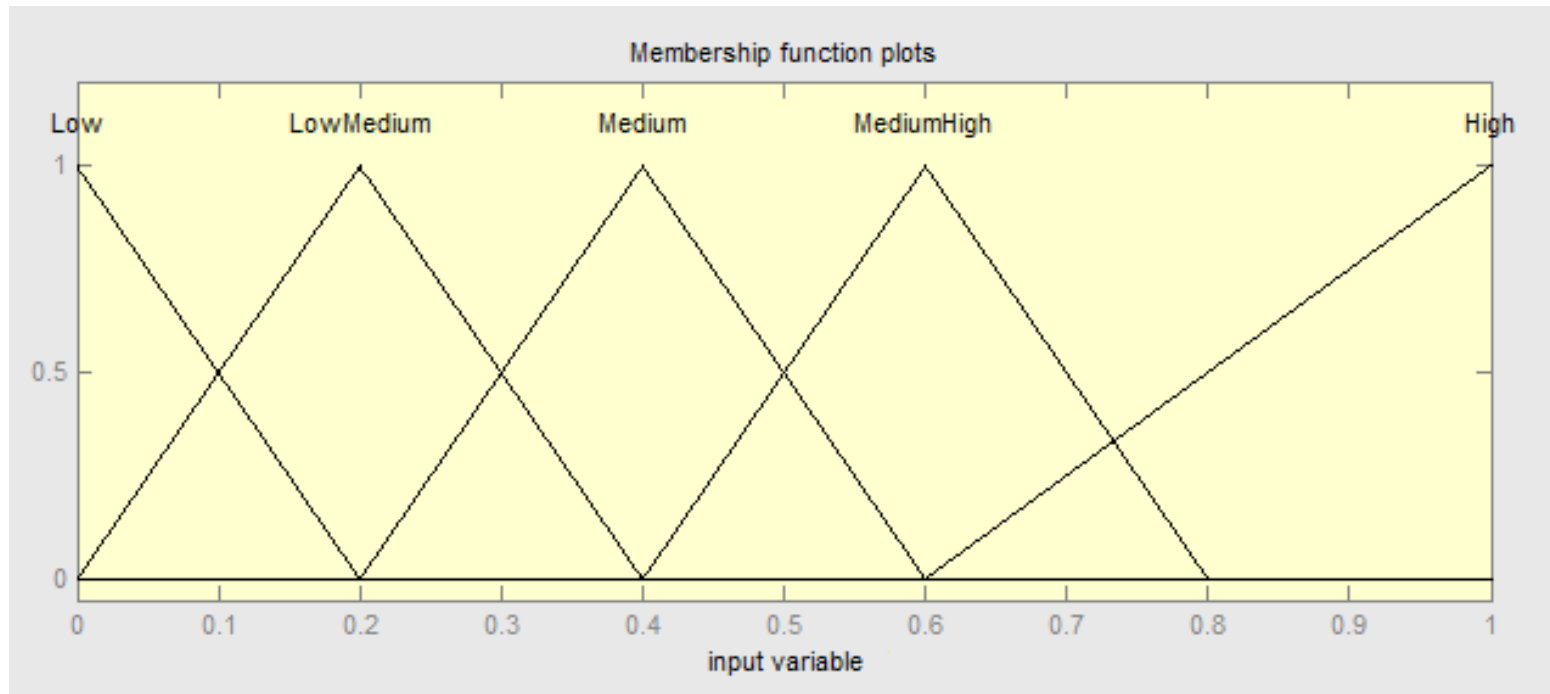


Parametre inferenčného systému FK

Počet vstupných premenných – 6

Maximálny počet klauzúl v pravidle – 6

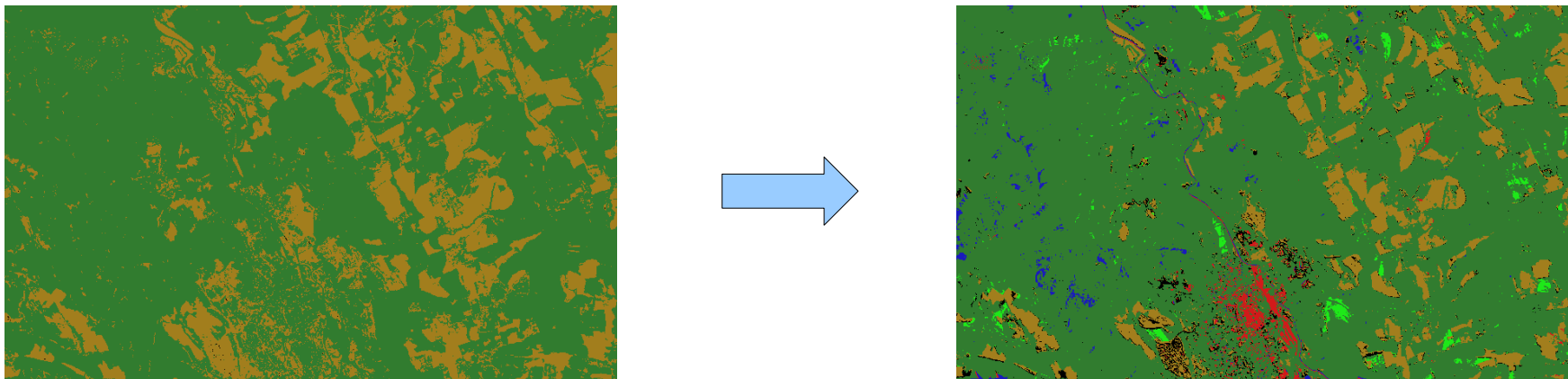
Každá premenná je fuzzifikovaná do 5 lingvistických termov



Inferencia prebieha podľa vzťahov uvedených na snímku 6.

Úpravy v pôvodnom návrhu

Dôsledok problémov s klasifikáciou nízko početných tried.



Vzťah určujúci istotu pravidla po úprave zohľadňuje váhy vzoriek:

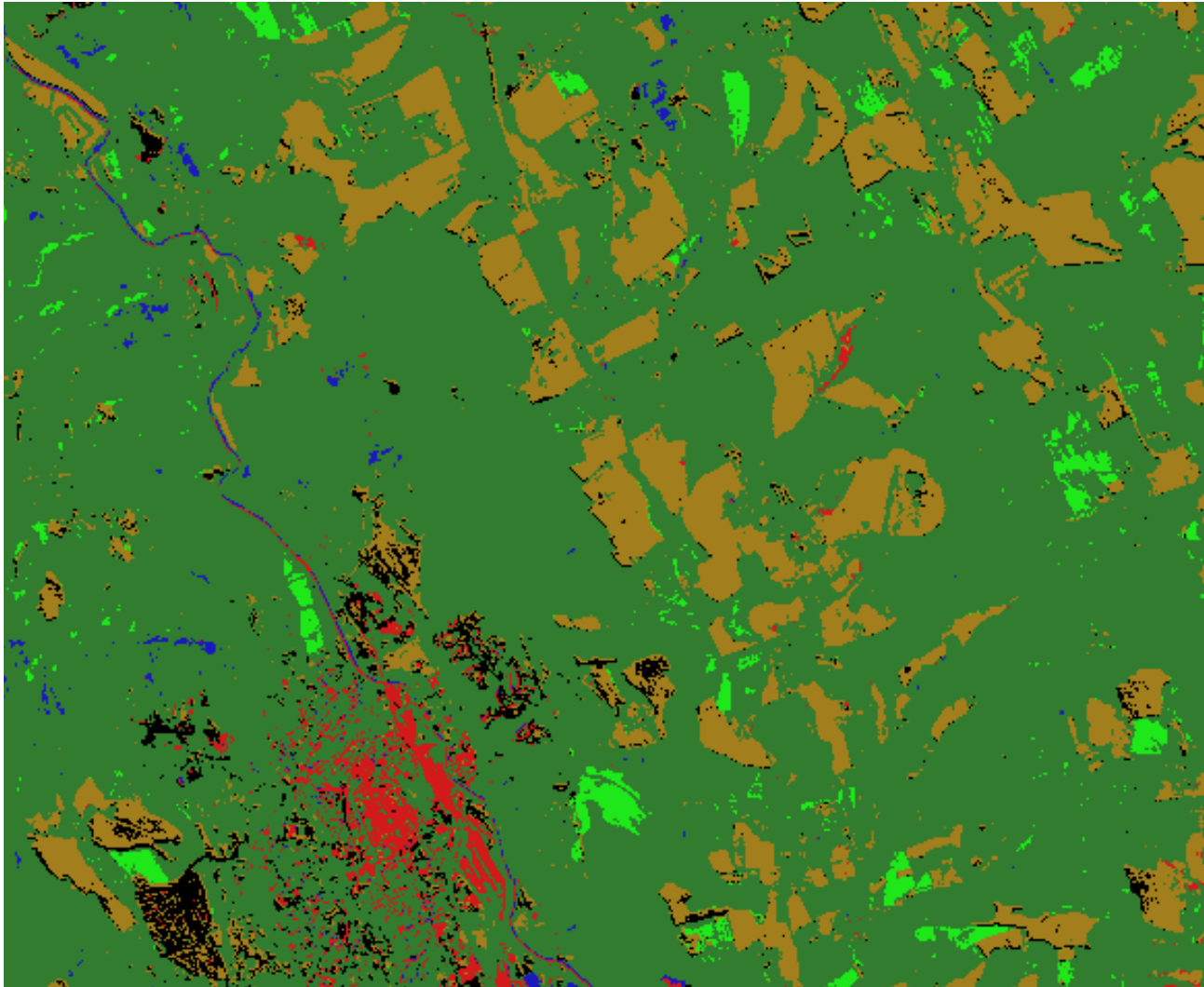
$$r^k = \sum_{p|c^p=C^k} w^p \mu^k(x^p) / \sum_{p=1}^Q w^p \mu^k(x^p)$$

Upravená istota bola zahrnutá do výpočtu vhodnosti („fitness”):

$$f_1 = \text{Cons} \cdot \text{Cov} \cdot r^k$$

Zavedená možnosť reštartu GA pri nízkej hodnote istoty pravidla.

Experiment č.1



Parametre učenia:

Veľkosť populácie: 50

Počet generácií: 10000

Mutácia: 20 %

Elitárstvo: 5 %

Poč. gen. na pravidlo: 20

Min. istota pravidla: 60 %

Poč. gen. na reštart GA: 10

Výsledky učenia:

Počet pravidiel: 24

Počet správne klasifikovaných vzoriek: 2528 z 3164

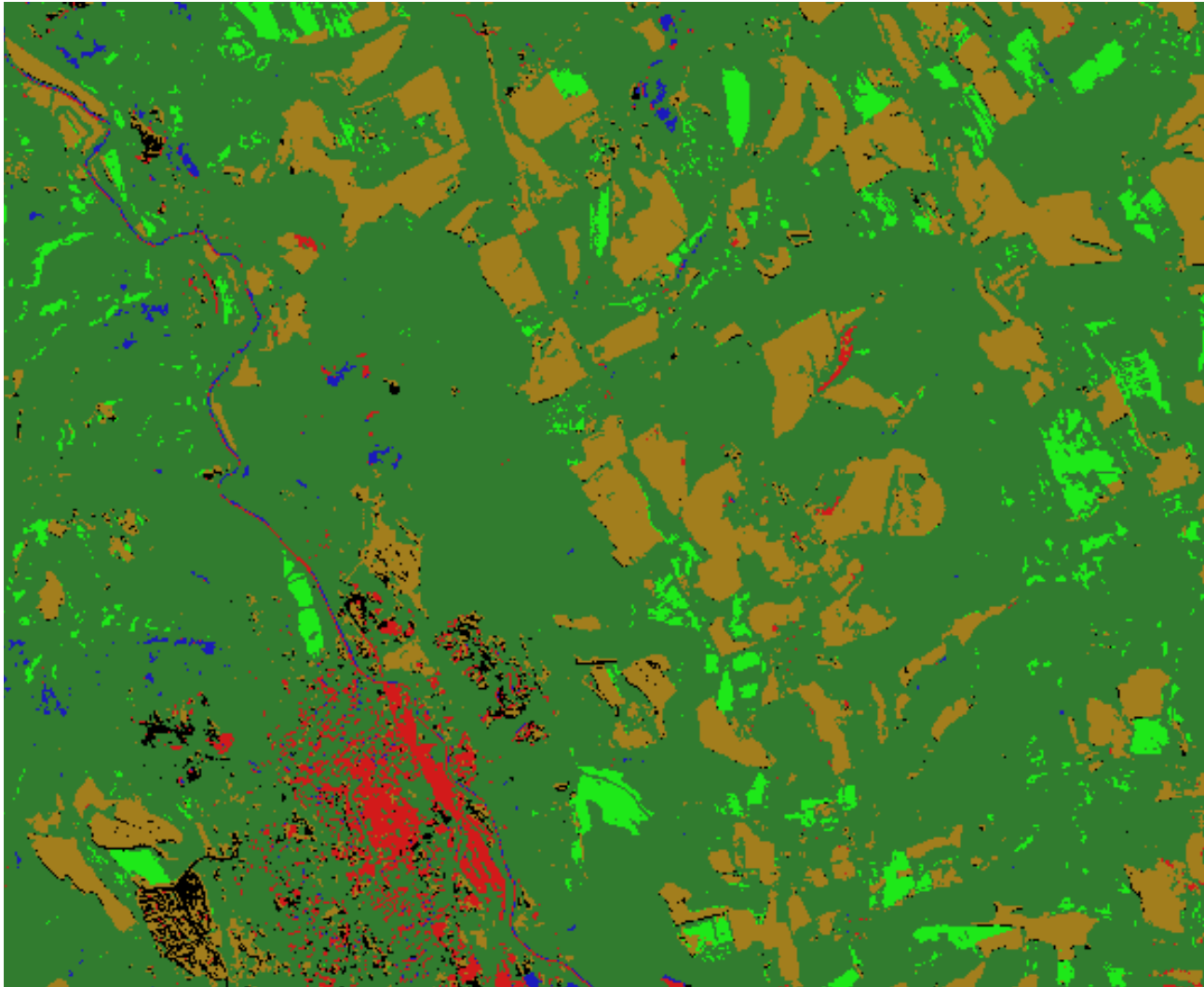
Presnosť klasifikácie: 79,90 %.

Výsledky testovania:

Počet správne klasifikovaných vzoriek: 2739 z 3167

Presnosť klasifikácie: 86,49 %

Experiment č.2



Parametre učenia:

Veľkosť populácie: 100

Počet generácií: 1000

Mutácia: 20 %

Elitárstvo: 10 %

Poč. gen. na pravidlo: 20

Min. istota pravidla: 50 %

Poč. gen. na reštart GA: 100

Výsledky učenia:

Počet pravidiel: 50

Počet správne klasifikovaných vzoriek: 2413 z 3164

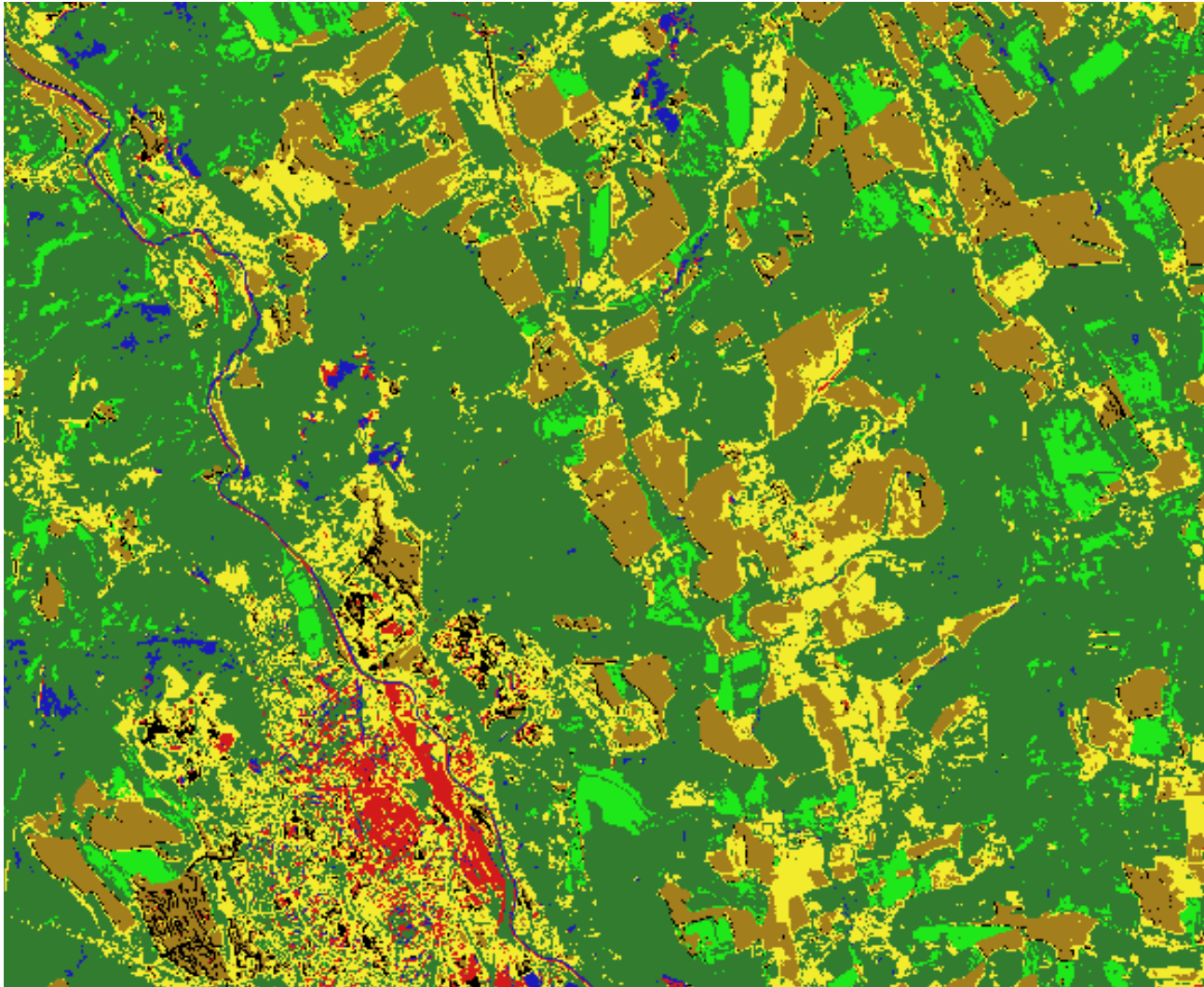
Presnosť klasifikácie: 76,26 %.

Výsledky testovania:

Počet správne klasifikovaných vzoriek: 2637 z 3167

Presnosť klasifikácie: 83,26 %

Experiment č.3



Parametre učenia:

Veľkosť populácie: 100

Počet generácií: 10000

Mutácia: 20 %

Elitárstvo: 10 %

Poč. gen. na pravidlo: 20

Min. istota pravidla: 50 %

Poč. gen. na reštart GA: 100

Výsledky učenia:

Počet pravidiel: 50

Počet správne klasifikovaných vzoriek: 2466 z 3164

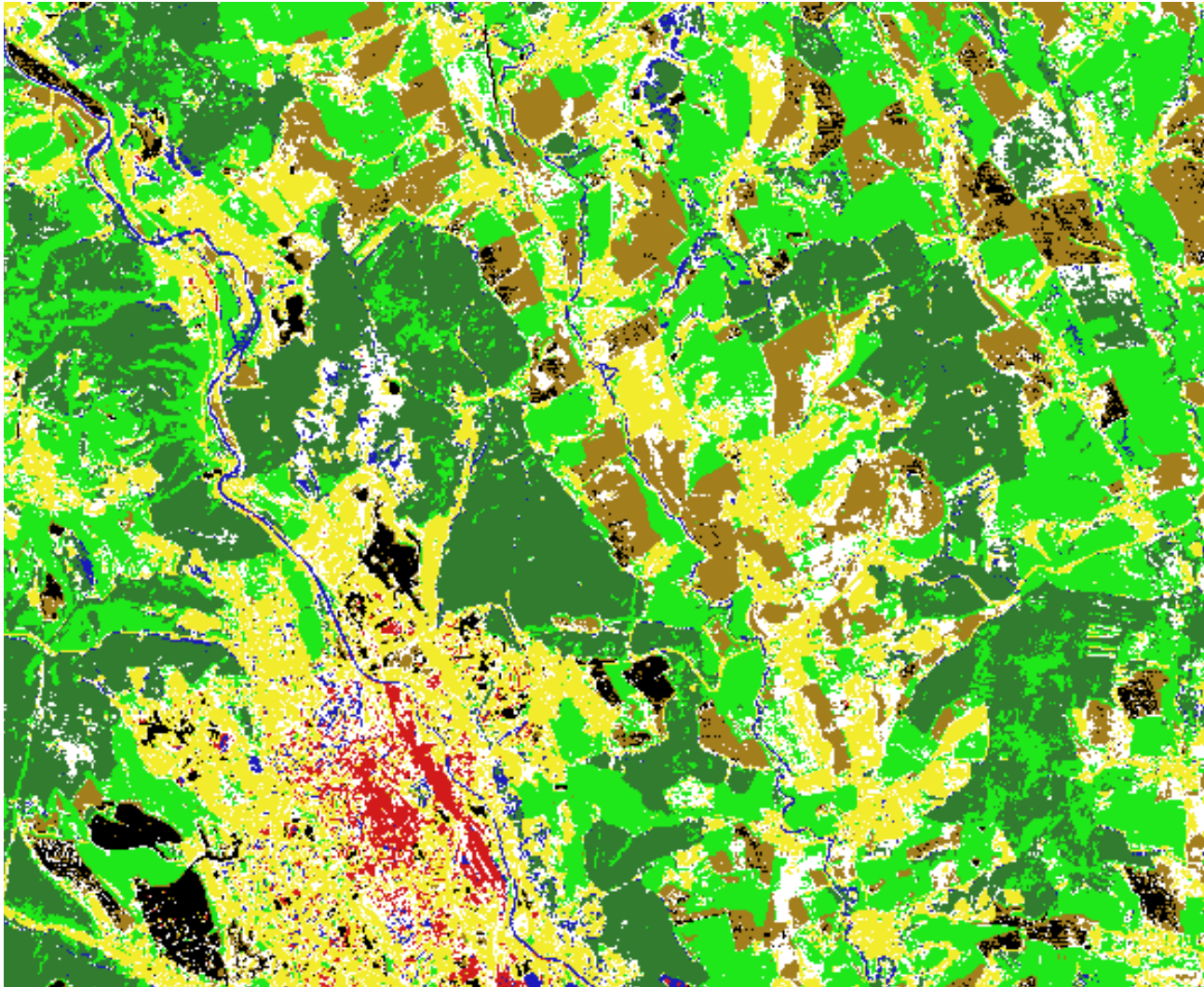
Presnosť klasifikácie: 77,94 %.

Výsledky testovania:

Počet správne klasifikovaných vzoriek: 2650 z 3167

Presnosť klasifikácie: 83,68 %

Porovnanie s klasifikáciou NS



Parametre siete:

Typ siete: Dopredná NS
2 skryté vrstvy: 8x6 neurónov
Typ učenia: Backpropagation
Počet cyklov učenia: 10000

Výsledky testovania:

Presnosť klasifikácie: 83,54 %

Záver

Nevýhody:

- Zložitá implementácia
- Nízka citlivosť na málo početné triedy
- Ťažkosti s hľadaním optimálnych parametrov učenia
- Vyššia časová náročnosť učenia ako pri NS

Výhody:

- Automatizácia učenia
- Jednoduchý prístup k uloženým znalostiam (pravidlám)
- Možnosti ďalšej optimalizácie pravidiel
- Väčšinu nevýhod je možné do budúcnosti odstrániť

Ďakujem za pozornosť!

??? Otázky ???