

Abstrakt

Diplomová práce se zabývá optimalizací trénovacího a validačního datasetu pro řízenou klasifikaci dat v DPZ. V rámci řešení práce jsou v území lesně-luční krajiny v Podkrkonoší prováděny pro dva klasifikační algoritmy (Maximum Likelihood – MLC a Support Vector Machine – SVM) experimenty s trénovacími a validačními daty. Práce vychází z předpokladu, že pro dosažení maximální přesnosti klasifikace je ideální podíl 1/3 trénovacích a 2/3 validačních dat (Foody, 2009). Další hypotézou práce byl předpoklad, že v případě klasifikace pomocí algoritmu SVM je pro dosažení stejné/podobné přesnosti klasifikace potřeba nižší počet trénovacích bodů než v případě klasifikačního algoritmu Maximum Likelihood (Foody, 2004). Cílem práce bylo testovat vliv podílu/množství trénovacích a validačních dat na přesnost klasifikace multispektrálních dat senzoru Sentinel-2A s využitím algoritmu Maximum Likelihood. Nejvyšší celkové přesnosti při využití klasifikačního algoritmu Maximum Likelihood bylo dosaženo pro podíl 375 trénovacích a 625 validačních bodů. Celková přesnost pro tento podíl byla 72,88 %. Teorie Foodyho (2009), že pro dosažení nejvyšší přesnosti klasifikace je ideální podíl 1/3 trénovacích a 2/3 validačních dat potvrzují výsledky hodnocení celkové přesnosti a Kappa koeficientu pro Maximum Likelihood. Avšak výsledné uživatelské a zpracovatelské přesnosti pro jednotlivé třídy nedosáhly v případě tohoto podílu nejvyšších hodnot. Také se ukázalo, že změna velikosti validačního datasetu při zachování stabilní velikosti trénovacího datasetu má vliv na stabilitu výsledku hodnocení přesnosti klasifikace metodou Maximum Likelihood. Výsledek hodnocení celkové přesnosti pro algoritmus SVM pro tento podíl 375 trénovacích a 625 validačních bodů byl 79,09 %. V případě, že se počet trénovacích bodů snížil na 50 byla celková přesnost klasifikace 76,49 %. Předpoklad, že pomocí algoritmu SVM je pro dosažení stejné/podobné přesnosti klasifikace potřeba nižší počet trénovacích bodů než v případě klasifikačního algoritmu Maximum Likelihood, byl pro tento dataset potvrzen.

Klíčová slova: trénovací dataset, validační dataset, hodnocení přesnosti klasifikace, Maximum Likelihood, Support Vector Machine