

# Používateľská príručka k skriptom na filtrovanie snímkov

Táto časť príručky poskytuje pokyny na použitie štyroch Python skriptov navrhnutých na filtrovanie datasetov histopatologických snímkov na základe farby, rozmazania a počtu buniek. Skripty obsahujú GUI pre jednoduché použitie a podporujú režimy filtrovania aj optimalizácie.

## Systémové požiadavky

- **Verzia Pythonu:** Python 3.8 alebo vyšší
- **Hardvér:** Dostatočný priestor na disku pre vstupné snímky a výstupné priečinky.
- **Vstupné súbory:** Snímky vo formátoch PNG alebo JPEG.

## Inštalácia a príprava

Stiahnite a nainštalujte Python 3.8+ z [python.org](https://python.org). Uistite sa, že je nainštalovaný `pip`. Otvorte terminál a spustite:

```
pip install opencv-python numpy matplotlib pandas imutils scikit-learn Pillow webcolors
```

Uložte všetky skripty do jedného priečinka, pretože `full_filter.py` závisí od ostatných troch. Umiestnite svoje snímky do zvoleného priečinka. Ak plánujete optimalizovať niektorú z metód, uistite sa, že názvy súborov sú popisné - obsahujú *good* pre vhodné a *bad* pre nevhodné snímky. V prípade filtrácie nie sú na názvy súborov nijaké požiadavky, nemusíte ich nijak upravovať.

### `color__filter__script.py`

Spracuje snímky na základe dominantných farieb (ružová, biela, ružovo-biela, fialová):  
**Filtrovanie (0):** Presúva snímky do podpriečinkov na základe vybraných farieb (predvolene

sú vybraté všetky farby okrem fialovej) a ich farebných hraničných hodnôt (predvolené hodnoty v zátvorke):

- *good* = vhodné,
- *pink* = príliš ružové, percento  $\geq$  hraničná hodnota (51)
- *white* = príliš biele, percento  $\geq$  hraničná hodnota (64)
- *pink\_white* = príliš ružovo-biele, percento  $\geq$  hraničná hodnota (36)
- *not\_enough\_purple* = nedostatočne fialové, percento  $<$  hraničná hodnota (0)

**Optimalizácia (1):** Analyzuje výkon (senzitivitu, špecificitu, celkovú presnosť) pre hraničné hodnoty pre vybranú farbu a ukladá výsledky do súboru *color\_analysis.csv*. Rozsah testovaných hraničných hodnôt je automaticky 0 až 100 a nie je nastaviteľný. **Analýza jedného snímku (2):** Analyzuje rozloženie farieb jedného konkrétneho snímku. Výstup v termináli uvádza farebné skupiny, percentá a kategórie. Zobrazí sa spracovaný snímok a koláčový graf rozloženia farieb.

#### Vstupy GUI:

- **Typ spustenia (Run type):** Vyberte 0 (Filtrovanie), 1 (Optimalizácia) alebo 2 (Analýza jedného snímku).
- **Vstupný priečink (Input folder):** Cesta k priečinku so snímkami.
- **Možnosti farieb (Select colors)** pre filtrovanie: Zaškrŕavacie políčka pre ružovú (Pink), ružovo-bielu (Pink\_white), bielu (White) a fialovú (Purple). Ak je farba vybraná, bude sa podľa nej filtrovať obsah vstupného priečinka.
- **Hraničné hodnoty (Threshold)** pre filtrovanie: Kladné číselné hodnoty pre každú vybranú farbu.
- **Výber farby (Select colors)** pre optimalizáciu: Prepínač pre jednu farbu, ktorej hraničnú hodnotu chcete optimalizovať.
- **Cesta k snímku (Image path)** pre analýzu jedného snímku: Cesta k jednému snímku.

## **blur\_filter\_script.py**

Spracuje snímky na základe úrovne rozmazania: **Filtrovanie (0)**: Presúva snímky do priečinkov *good* (nerozmazané) alebo *bad* (rozmazané) na základe hraničnej hodnoty rozmazania (predvolene 215). **Optimalizácia (1)**: Analyzuje výkon (senzitivitu, špecificitu, celkovú presnosť) pre zadaný rozsah hraničných hodnôt (predvolene 50 až 500, s krokom 5) a ukladá výsledky do *blur\_analysis.csv*.

### **Vstupy GUI:**

- **Typ spustenia (Run type)**: Vyberte 0 (Filtrovanie) alebo 1 (Optimalizácia).
- **Vstupný priečinok (Input folder)**: Cesta k priečinku so snímkami.
- **Prah rozmazania (Blur threshold)** pre filtrovanie: Kladné číslo.
- **Rozsah prahov (Threshold range)** pre optimalizáciu: Kladné číslo pre začiatok (Start), koniec (End) a krok (Step). Začiatok musí byť menší ako koniec.

## **cell\_filter\_script.py**

Filtruje snímky na základe počtu detegovaných buniek **Filtrovanie (0)**: Presúva snímky do priečinkov *good* (počet buniek  $\geq$  hraničná hodnota) alebo *bad* (počet buniek  $<$  hraničná hodnota, predvolená 25). **Optimalizácia (1)**: Analyzuje výkon (senzitivitu, špecificitu, celkovú presnosť) pre rozsah hraničných hodnôt (predvolene 0 až 50 s krokom 1) a ukladá výsledky do *cell\_count\_analysis.csv*.

### **Vstupy GUI:**

- **Typ spustenia (Run type)**: Vyberte 0 (Filtrovanie) alebo 1 (Optimalizácia).
- **Vstupný priečinok (Input folder)**: Cesta k priečinku so snímkami.
- **Hraničná hodnota (Threshold)** pre filtrovanie: Počet buniek.
- **Rozsah prahov (Threshold range)** pre optimalizáciu: Kladné číslo pre začiatok (Start), koniec (End) a krok (Step). Začiatok musí byť menší ako koniec.

## **full\_filter.py**

Kombinuje vyššie uvedené skripty do sekvenčného procesu s nasledujúcimi krokmi:

1. Farebný filter (ružová, biela, ružovo-biela).
2. Filter počtu buniek na priechkoch ružovej a ružovo-bielej.
3. Filter rozmazania na kombinovaných dobrých snímkach.
4. Filter počtu buniek na nerozmazaných snímkach.
5. Farebný filter (fialová) na zlých snímkach.

### **Vstupy GUI:**

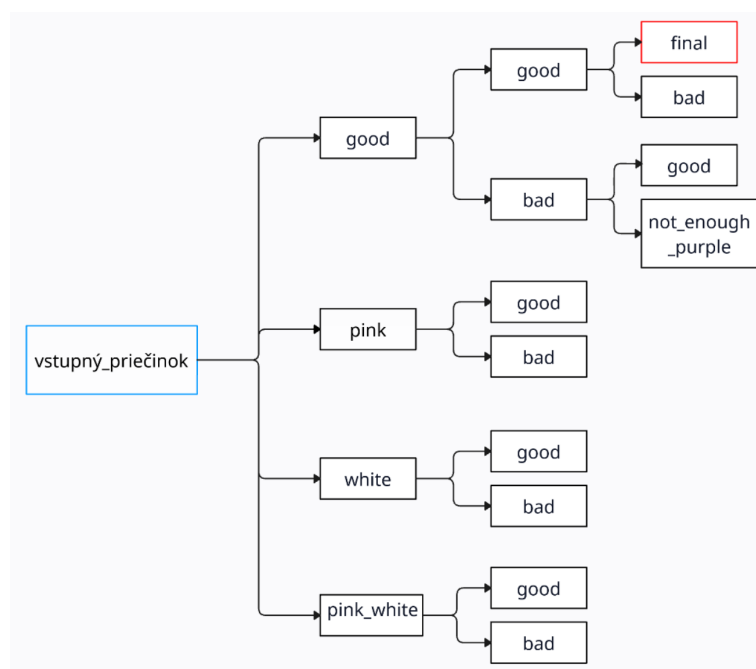
- **Vstupný priečinok (Input folder):** Cesta k počiatočnému priečinku so snímkami (napr. `./dataset`).
- **Hraničné hodnoty:**
  - ružovej (predvolená hodnota 51)
  - bielej (predvolená hodnota 36)
  - ružovo-bielej (predvolená hodnota 64)
  - počtu buniek (predvolená hodnota 25)
  - rozmazania (predvolená hodnota 215)
  - fialovej (predvolená hodnota 62)

Finálne filtrované snímky sú v `vstupný_priečinok/good/good/final`. Príklad štruktúry priečinkov pre `full_filter.py` je na obrázku 1.

## **Spustenie a výstupy**

Prejdite do priečinka so skriptami v termináli. Spustite skript príkazom:

```
python názov_skriptu.py
```



Obr. 1: Hierarchia priečinkov po filtrácii skriptom `full_filter.py`

Po spustení skriptu sa otvorí okno GUI. Zadať požadované vstupy (cesty k priečinkom, hraničné hodnoty atď.). Použite tlačidlo *Browse* (prehľadávať) na výber priečinkov alebo súborov. Kliknite na *Run analysis* (spustiť analýzu) alebo *Run filtering* (spustiť filtrovanie) na začatie spracovania. GUI sa zatvorí a spracovanie začne v termináli. V termináli sa zobrazia správy o priebehu. Pre režim optimalizácie sa zobrazuje priebeh pre každú hraničnú hodnotu.

**Režim filtrovania:** Snímky sú presunuté do podpriečinkov (napr. *pink*, *pink\_white*, *bad*), ktoré obsahujú snímky vylúčené v každom kroku.

**Režim optimalizácie:** Vytvorí sa CSV súbor vo vstupnom priečinku. Stĺpce zahŕňajú *Threshold* (hraničnú hodnotu), *Sensitivity* (senzitivitu), *Specificity* (špecificitu) a *Accuracy* (celková presnosť). Na záver sa zobrazí graf vývoja metrík.

## Riešenie problémov a testovanie

V prípade nevhodného vstupu (napr. záporné číslo, začiatok rozsahu väčší ako jeho koniec) GUI informuje o chybe a nespustí proces. V prípade nevyplnenia povinného poľa (napr. vstupný priečinok) GUI takisto informuje o chybe a nespustí proces. V prípade, že sa žiadne snímky

nespracujú, overte či vstupný priečinok obsahuje súbory vo formáte PNG alebo JPEG.

Otestujte na malom datasete, aby ste overili nastavenia pred spracovaním veľkých datasetov. Pred spustením zálohujte snímky, pretože skripty presúvajú súbory. Optimalizácia a filtrácia môže podľa rozsahu testovaných hodnôt a počtu testovaných snímok trvať aj niekoľko hodín.