

Co může databázím přinést automatické rozpoznávání rostlin?

Ondřej Vild, Ondřej Veselý, Karel Chobot

Analýza obrazu pomocí hlubokých neuronových sítí se postupně dostává z odborných konferencí do běžného života. Mezi její aplikace patří i rozpoznávání rostlin z fotografií. Aplikace fungují jednoduše: požadovanou rostlinu vyfotíte na smartphone a za pár oka-

mžků dostanete název rostliny. Jednou z nejúspěšnějších aplikací v oboru je aplikace Plant.id brněnské firmy FlowerChecker. Tuto technologii je ale také možné využít pro identifikaci většího počtu fotografií a celých jejich databází.

Databáze fotografií v NDOP

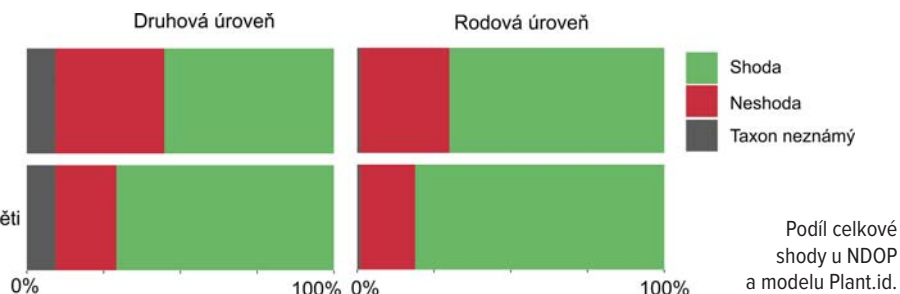
Jednu takovou databázi vlastní a spravuje Agentura ochrany přírody a krajiny ČR. Je součástí Nálezové databáze ochrany přírody, která obsahuje data o výskytu rostlin, hub a živočichů v České republice. Záznamy v ní pocházejí nejen od expertů, ale skrze veřejně dostupnou aplikaci BioLog do ní mohou přispívat i amatérští přírodovědci. Tyto záznamy jsou méně spolehlivé, a proto je zde na místě jejich ověření experty. Tato kontrola je ale časově poměrně náročná a využití technologie Plant.id se zde nabízí. Zajímalo nás, jestli by bylo možné pomocí této technologie usnadnit kontrolu záznamů.

Použití modelu Plant.id

Pro analýzu jsme použili všechny dostupné záznamy s fotografiemi cévnatých rostlin, pořízené mezi lety 2014 a 2020. Celkem se jednalo o 10 131 záznamů s 1552 unikátními taxony, nejčastěji v druhové úrovni. Všechny fotografie jsme následně analyzovali modelem hluboké konvoluční sítě Plant.id. Výsledkem pro každou fotografii byl seznam nejpravděpodobnějších druhů.

Porovnání autorů záznamu a modelu

To nám umožnilo porovnat, do jaké míry se data z NDOP shodují s výsledky modelu. Podívali jsme se na procentuální shodu na rodové a druhové úrovni. Kromě shody u prvního výsledku modelu jsme také zaznamenávali, jestli existuje shoda v nějakém z prvních pěti výsledků modelu. Při vyhodnocení jsme zohlednili synonyma. Výsledky ukázaly, že shoda na druhové úrovni u prvního výsledku byla u 55 % fotografií a na rodové úrovni u 70 % fotografií. Shoda v jednom z pěti výsledků pak byla na druhové úrovni



u 71 % fotografií a na rodové úrovni u 81 % fotografií. Vzhledem k tomu, že model Plant.id „nezná“ všechny taxony z NDOP, nebyla shoda možná v 9 % případech u druhové úrovně a v 1 % případech u rodové úrovně. Jednalo se především o vzácnější a kritické taxony. Je tato míra shody „úspěchem“? Při podobném přístupu, na fotografiích rostlin britské flóry, se Plant.id trefilo do druhu v prvním výsledku u téměř 70 % případů (Jones 2020). Tento 15% rozdíl může být způsoben rozdílnou kvalitou fotografií, ale také právě vyšší jistotou identifikace v britské studii, pokud připustíme, že v databázi fotografií v rámci NDOP mohou být pozorováni s chybným určením.

Revize fotografií

Abychom zjistili, jestli k takovým k neshodám dochází, potažmo v jakých případech, seřadili jsme si záznamy podle míry jistoty modelu. Prvních 20 záznamů následně Ondřej Vild a Ondřej Popelka revidovali na základě fotografií a znalosti rozšíření druhů. Ve čtyřech případech jsme dali za pravdu autorovi záznamu a ve 14 případech Plant.ID modelu. Ve dvou případech se na fotografii vyskytovaly oba druhy. Revize ukázala, že pokud si je model jistý, má často opravdu pravdu. Dá se nicméně předpokládat, že s klesajícím sebevědomím modelu se bude jazýček na vahách přesouvat ve prospěch autora záznamu. To bude nejspíše platit u méně častých a kritických taxonů, dále u méně kvalitních fotografií a nekvetoucích jedinců.

Názor autora a modelu na pojmenování druhu u prvních 20 záznamů, kde si byl model Plant.id nejistější.

Číslo fotografie	Autor	Model Plant.id	Jistota modelu	Vítěz revize
1	<i>Ajuga reptans</i>	<i>Prunella vulgaris</i>	0,997	Plant.id
2	<i>Trifolium pratense</i>	<i>Trifolium repens</i>	0,997	Plant.id
3	<i>Geum rivale</i>	<i>Geranium phaeum</i>	0,996	Plant.id
4	<i>Cephalanthera damasonium</i>	<i>Platanthera bifolia</i>	0,996	Plant.id
5	<i>Primula veris</i>	<i>Primula elatior</i>	0,995	Plant.id
6	<i>Lysimachia nemorum</i>	<i>Lysimachia nummularia</i>	0,995	Plant.id
7	<i>Matricaria recutita</i>	<i>Chelidonium majus</i>	0,994	Plant.id
8	<i>Lysimachia vulgaris</i>	<i>Lysimachia punctata</i>	0,994	Plant.id
9	<i>Tephrosia crista</i>	<i>Horminum pyrenaicum</i>	0,991	NDOP
10	<i>Chrysosplenium alternifolium</i>	<i>Chrysosplenium oppositifolium</i>	0,991	NDOP
11	<i>Ajuga reptans</i>	<i>Ajuga genevensis</i>	0,99	Plant.id
12	<i>Platanthera chlorantha</i>	<i>Plantago lanceolata</i>	0,989	NDOP i Plant.id
13	<i>Lamium maculatum</i>	<i>Ballota nigra</i>	0,988	Plant.id
14	<i>Lamium purpureum</i>	<i>Lamium maculatum</i>	0,987	Plant.id
15	<i>Dactylorhiza fuchsii</i>	<i>Dactylorhiza maculata</i>	0,987	NDOP
16	<i>Pulmonaria officinalis</i>	<i>Symphytum officinale</i>	0,987	Plant.id
17	<i>Ajuga reptans</i>	<i>Ajuga pyramidalis</i>	0,987	Plant.id
18	<i>Chrysosplenium alternifolium</i>	<i>Chrysosplenium oppositifolium</i>	0,986	NDOP
19	<i>Lathyrus niger</i>	<i>Vicia sepium</i>	0,985	Plant.id
20	<i>Ailanthus altissima</i>	<i>Rhus typhina</i>	0,984	NDOP i Plant.id


Ukázka výsledku modelu: jednotlivé taxony seřazené podle pravděpodobnosti.



za pár sekund

změnit jazyk rostliny (cs)

smazat

tato rostlina vypadá zdravě 

Vicia sepium

vikev plotní

 Google  Podrobnosti

88.6%
potvrdit



Vicia sativa

 Google  Podrobnosti

4.0%
potvrdit



Vicia cracca

vikev ptačí

 Google  Podrobnosti

2.0%
potvrdit



Závěr

Výsledky ukazují, že technologie rozpoznávání rostlin z fotografií může usnadnit revizi záznamů v rozsáhlejších databázích, jako například NDOP. AOPK ČR uvažuje i o tom, že tento nástroj začlení do aplikace BioLog. Tím by mohl být uživatel upozorněn na možnou záměnu ještě před vložením záznamu, v důsledku čehož by se celá databáze stala spolehlivější. Mohlo by to též zvýšit ochotu záznamy o výskytu druhů do databáze vkládat, zvláště u opatrnějších uživatelů.

Díky aplikaci Insect.id vydané v lednu 2023 bude možné podobně vyhodnotit i bezobratlé živočichy a postupně i různé skupiny obratlovců. ■

Odkazy:

Jones, H. G. (2020). What plant is that? Tests of automated image recognition apps for plant identification on plants from the British flora. *AoB PLANTS*, 12(6), 1–9. <https://doi.org/10.1093/aobpla/plaa052>

Článek vznikl také s pomocí prostředků podporujících dlouhodobý koncepční rozvoj výzkumné organizace – RVO 67985939.