**Prepoznavanje zdravih i bolesnih listova biljaka korišćenjem**

**konvolucionih neuronskih mreža (CNN)**

**i metode potpornih vektora (SVM)**

Cilj projekta je razviti dva modela za binarnu klasifikaciju listova biljaka:

* **CNN model** – direktno klasifikuje slike listova kao „Healthy“ ili „Diseased“ koristeći duboke konvolucione mreže.
* **SVM model** – klasifikuje slike koristeći karakteristike ekstraktovane iz CNN-a

Problem se rešava na osnovu vizuelnih podataka, slika listova biljaka, i ima cilj da precizno odredi zdravlje biljaka, čime se olakšava preventivno delovanje u poljoprivredi. Kombinacija CNN-a i SVM-a omogućava upoređivanje performansi modernih dubokih mreža i tradicionalnih klasifikatora, čime se može postići optimalna tačnost i robustnost.

**1. Motivacija**

Rana detekcija bolesti biljaka značajno smanjuje gubitke u poljoprivredi i povećava prinos. Automatski sistemi zasnovani na mašinskom učenju omogućavaju brzu i pouzdanu analizu velikog broja biljaka.

Korišćenje CNN-a i SVM-a pruža robustan pristup koji može da podrži poljoprivrednike u donošenju odluka i primeni preventivnih mera.

**2. Skup podataka**

**PlantVillage dataset**

* **Atributi**: RGB slike listova biljaka (128x128 piksela)
* **Ciljno obeležje**: Healthy (1) ili Diseased (0)
* **Broj instanci i raspodela po klasama**:
  + - Healthy: 3069
    - Diseased: 2997
    - Ukupno: 6066 instanci
* **Link**: <https://www.kaggle.com/datasets/emmarex/plantdisease/data>

Svaka instanca predstavlja pojedinačnu sliku listova biljke koja se koristi za učenje modela.

**3. Pretprocesiranje podataka**

Sve slike se menjaju na veličinu od 128×128 piksela i normalizuju u opseg vrednosti [0,1] kako bi model lakše učio.

Podaci se zatim dele na train, validation i test skupove u odnosu 70%, 15% i 15%, pri čemu se podaci mešaju kako bi model učio karakteristike, a ne redosled slika.

Za SVM model koristiće se feature-i ekstraktovani iz CNN Flatten sloja kao ulazni vektor.

**4. Metodologija**

**4.1 CNN metod**

CNN model direktno uči vizuelne karakteristike slika i klasifikuje ih kao Healthy ili Diseased.

Ulaz modela predstavlja **RGB slika** dimenzija 128×128 piksela.

Tokom treninga koristiće se funkcija gubitka **binary\_crossentropy**, dok će hiperparametri, kao što su broj epoha i veličina batch-a, biti podešavani pomoću validacionog skupa.

Performanse modela će se pratiti kroz metrike **accuracy** i **loss** po epohi, kako bi se identifikovalo eventualno pretreniravanje (overfitting) ili nedovoljno učenje (underfitting).

**4.2 SVM metod**

SVM klasifikuje slike na osnovu karakteristika ekstraktovanih iz CNN modela.

Dobijeni feature-i se standardizuju i dimenzionalnost se smanjuje radi efikasnijeg treniranja primenom **PCA**.

Model će koristiti **sklearn.svm.SVC** sa **RBF** ili **linear** **kernelom**, što omogućava da, pored same klase, model vrati i verovatnoću predikcije.

Hiperparametri, kao što su **C**, **gamma** i tip kernela, biće podešavani pomoću **GridSearchCV**, dok će se za odabir najbolje kombinacije koristiti **k-fold cross-validation**.

**5. Način evaluacije**

Performanse oba modela biće evaluirane na test skupu korišćenjem **metrike** **tačnosti** (accuracy), **preciznosti** (precision), **odziva** (recall) i **F1-score**, dok će **confusion matrix** omogućiti detaljniju analizu grešaka i identifikaciju potencijalnih problema u klasifikaciji.

Na kraju se upoređuju performanse SVM-a i CNN-a kako bi se odabrao optimalan pristup za problem klasifikacije listova.

**6. Tehnologije**

* **Programski jezik**: Python 3
* **Duboko učenje (CNN)**: TensorFlow/Keras
* **Mašinsko učenje (SVM)**: scikit-learn (SVC, GridSearchCV)
* **Analiza podataka**: NumPy, Pandas
* **Vizualizacija**: Matplotlib, Seaborn
* **Okruženje**: Google Colab (GPU podrška)
* **Druge alatke**: os, shutil, random

**7. Relevantna literatura**

* <https://www.youtube.com/watch?v=jztwpsIzEGc>
* <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1412988/full>
* <https://openagriculturejournal.com/VOLUME/18/ELOCATOR/e18743315305194/FULLTEXT>