



UNIVERZITET CRNE GORE

PRIRODNO MATEMATIČKI FAKULTET

## **SEMINARSKI RAD**

IDENTIFIKACIJA BOLESTI LIŠĆA ČILJJA KORIŠĆENJEM  
IZJEDNAČAVANJA HISTOGRAMA SA OGRANIČENIM KONTRASTOM I  
GRUPISANJA K-SREDNJIH VRIJEDNOSTI

Autor:

*Igor Banović*

Profesor:

*Igor Jovančić*

## **Sadrzaj**

1. Uvod .....	3
2. Problem i motivacija .....	4
3. Cilj i doprinos rada .....	6
4. Pregled datog istraživačkog rada .....	7
5. Metode korišćene u istraživačkom radu .....	9
6. Analiza rezultata .....	13
7. Ograničenja i moguća poboljšanja .....	15
8. Zaključak .....	16
9. Prijedlog projekta.....	17
Reference.....	19

## 1. Uvod

Bolesti biljaka predstavljaju jedan od najvećih izazova u savremenoj poljoprivredi, jer značajno utiču na prinos, kvalitet i ekonomsku isplativost proizvodnje. Biljke su stalno izložene napadima insekata, gljivičnih, virusnih i bakterijskih patogena, što dovodi do postepenog propadanja usjeva ukoliko se infekcije ne otkriju na vrijeme. Najranjivije su biljke kod kojih list predstavlja glavni pokazatelj bolesti, kao što je čili paprika, gdje su promjene na lišću najraniji znak bolesti.

Čili je biljka široke primjene u prehrambenoj industriji, pa bolesti poput pepelnice, antraknoze, plamenjače i bakterijske pjegavosti mogu izazvati značajne ekonomske gubitke. Tradicionalne metode otkrivanja bolesti, zasnovane na vizuelnoj procjeni stručnjaka, spore su, subjektivne i teško primjenjive na velikim poljoprivrednim površinama. Zbog toga se sve veća pažnja posvećuje automatizovanim metodama zasnovanim na obradi slike i mašinskom učenju koje će pružiti preciznu, brzu i pouzdanu analizu.

Razvoj mašinskog učenja i obrade slike omogućio je nove pristupe prepoznavanju biljnih bolesti na osnovu vizuelnih karakteristika lišća, kao što su boja, tekstura i oblik oboljelih područja. Takvi sistemi doprinose digitalizaciji poljoprivrede, smanjuju potrebu za ručnom inspekcijom, ubrzavaju proces dijagnostike i omogućavaju pravovremeno reagovanje. U ovom seminarском radu analiziran je istraživački rad *“Identification of chilli leaf disease using contrast limited histogram equalisation and k-means clustering”*, koji predlaže metod zasnovan na kombinaciji poboljšanja kontrasta (CLAHE), segmentacije (k-means) i klasifikaciju (SVM) radi pouzdane identifikacije najčešćih bolesti čilija.

## **2. Problem i motivacija**

Bolesti lišća čili paprike, poput antraknoze, bakterijske pjegavosti i plamenjače, direktno utiču na smanjenje kvaliteta i količine prinosa. Ove bolesti narušavaju proces fotosinteze, usporavaju rast biljke i u težim slučajevima, dovode do njenog potpunog propadanja. Gubici se odražavaju ne samo na ekonomiju proizvođača, već i na stabilnost tržišta i dostupnost proizvoda. Međutim, identifikacija bolesti u praksi predstavlja kompleksan i višeslojan izazov. Na vizuelno prepoznavanje simptoma utiču brojni faktori:

- različito osvjetljenje tokom fotografisanja,
- sličnost vizuelnih simptoma različitih bolesti,
- preklapanje ili oštećenja lišća koja nisu patološka,
- različite boje i tekture pozadine,
- promjene uzrokovane starošću lišća, štetočinama ili uslovima okoline.

Ovakvi faktori otežavaju pravilnu dijagnozu čak i iskusnim agronomima, a ručna inspekcija je spora i nepraktična na velikim površinama. Zbog toga je neophodno razviti pouzdane i automatizovane metode koje mogu tačno razlikovati zdravo i oboljelo lišće na osnovu vizuelnih karakteristika.

### **Najčešće bolesti lišća čilija**

Bolesti čili paprike imaju različite uzroke i simptome. Pepelnica se karakteriše bijelim brašnastim premazom i žućenjem lišća, pri čemu se bolest brzo širi pri visokoj vlazi. Antraknoza izaziva tamne, udubljene pjege koje se mogu proširiti na plod, dok plamenjača stvara žute mrlje i sivkaste prevlake na donjoj strani lista. Bakterijska pjegavost dovodi do sitnih tamnih pjega sa žutim prstenom, a u težim slučajevima izaziva opadanje lišća. Razumijevanje ovih simptoma važno je za razvoj sistema koji treba da uči vizuelne karakteristike i skladišti ih u modele klasifikacije.

Zdravo lišće



Antraknoza



Bakterijska pjegavost



Pepelnica



Plamenjača



### **3. Cilj i doprinos rada**

Cilj istraživanog rada jeste razvoj metode za automatsku identifikaciju bolesti čili lišća na osnovu slike, kombinovanjem:

- poboljšanja kontrasta putem CLAHE algoritma,
- segmentacije oboljelog područja pomoću k-means klasterovanja,
- klasifikacije vrsta bolesti primjenom SVM modela.

Doprinos rada ogleda se u:

- integriranju navedenih tehnika u jedinstven sistem,
- evaluaciji performansi na otvorenom Kaggle skupu podataka,
- postizanju tačnosti od preko 97%,
- pokazivanju da klasične metode, uz adekvatnu predobradu, mogu biti konkurentne savremenim pristupima u specifičnim zadacima.

### **Naučna relevantnost istraživanja**

Relevantnost istraživanja ogleda se u tome da klasične metode obrade slike i dalje imaju veliku primjenu u slučajevima gdje nedostaju veliki datasetovi potrebni za duboko učenje. Pored toga, kombinacija CLAHE, k-means segmentacije i SVM klasifikacije omogućava izgradnju modela koji je računski veoma efikasan i može se koristiti na mobilnim uređajima. Ova karakteristika čini metod pogodnim za praktične aplikacije u poljoprivrednim sredinama gdje infrastruktura nije uvijek razvijena.

## **4. Pregled datog istraživačkog rada**

Istraživani rad bavi se razvojem metode za identifikaciju bolesti lišća čilija korišćenjem kombinacije tehnika obrade slike i mašinskog učenja. Autori organizuju istraživanje u nekoliko faza: prikupljanje podatka, predobrada slike, segmentacija, ekstrakcija karakteristika, klasifikacija i evaluacija modela.

U radu se koristi skup podataka preuzet sa platform Kaggle, koji sadrži slike tri kategorije lišća:

- zdravo lišće,
- lišće zaraženo antraknozom,
- lišće zaraženo bakterijskom pjegavošću.

Dataset je balansiran tako da svaka klasa ima približno jednak broj uzoraka, što poboljšava stabilnost treninga i izbjegava preferiranje pojedinih kategorija.

### **Eksperimentalno okruženje i evaluacija**

Model je testiran na različitim podjelama trening/test skupa, što omogućava procjenu stabilnosti klasifikatora. Rezultati ukazuju na to da model daje stabilne performanse pri različitim konfiguracijama treninga. Klasifikacija po pojedinačnim bolestima pokazuje tačnost višu od 96% za obje posmatrane bolesti.

### **Pozicioniranje rada u odnosu na dostupnu literaturu**

U istraživačkom radu dat je detaljan pregled dostupnih metoda za detekciju biljnih bolesti, uključujući:

- klasične tehnike obrade slike (thresholding, edge detection),
- metode mašinskog učenja (KNN, SVM, ANN),
- savremena CNN arhitektura (EfficientNet).

Autori naglašavaju da njihov pristup, iako jednostavniji od modernih dubokih modela, pruža visok stepen efikasnosti uz manju računarsku složenost. Zbog toga je sistem pogodan za primjenu na uređajima sa ograničenim resursima.

### **Ključne prednosti i utisci iz rada**

- Sistem je lak za implementaciju, jasan i računarski lagan.
- Kombinacija CLAHE + k-means + SVM omogućava visok nivo tačnosti na malom datasetu.
- Rezultati su jasni i ponovljivi zahvaljujući detaljnem opisu eksperimentalnog okruženja.

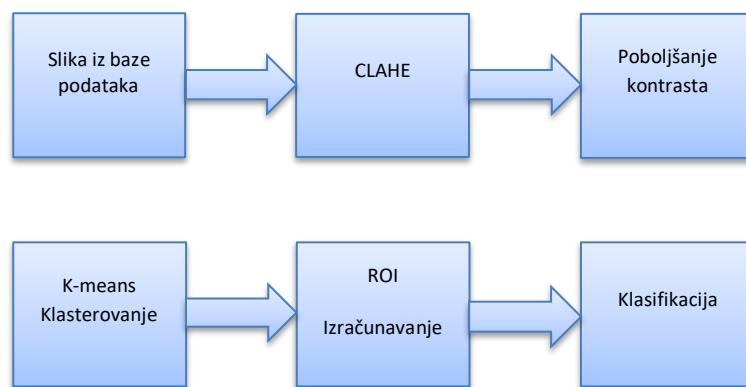
Međutim, istaknuto je da performanse na realnim poljoprivrednim slikama mogu odstupati od labaratoriskih rezultata, zbog čega se predlaže testiranje u terenskim uslovima i proširenje dataset-a.

Prije analize, autori primjenjuju CLAHE radi poboljšanja vizuelnog kvaliteta slika. Potom koriste k-means klasterovanje da izdvoje oboljele regije. Na kraju, SVM klasifikator donosi odluku o kojoj se vrsti bolesti radi.

U strukturi istraživačkog rada jasno se ističu faze: pregled literature, prikupljanje podataka, predobrada slika, segmentacija, klasifikacija i evaluacija rezultata. Autori detaljno opisuju matematičku pozadinu CLAHE metode, algoritamski tok k-means segmentacije i teorijsku i matematičku osnovu SVM-a. Pažnja je posvećena detaljnem opisu eksperimentalnog okruženja, uključujući karakteristike dataset-a, parametre treninga i kriterijume evaluacije, što obezbjeđuje veću jasnost i ponovljivost istraživanja.

## 5. Metode korišćene u istraživačkom radu

U istraživačkom radu primijenjen je kombinovani metod koji obuhvata tri ključne faze: poboljšanje kontrasta korišćenjem CLAHE algoritma, segmentaciju slike putem k-means klasterovanja i klasifikaciju tipa bolesti pomoću SVM modela. Ove tehnike zajedno čine efikasan postupak za identifikaciju bolesti lišća čilija, kako je prikazano u šematskom dijagramu metodologije.



### Obrada slike i poboljšanje kontrasta (CLAHE)

CLAHE (Contrast Limited Adaptive Histogram Equalisation) predstavlja prilagodljivu metodu za poboljšanje kontrasta slike. Za razliku od klasičnog histogram pristupa, CLAHE dijeli sliku na manje blokove i vrši lokalno izjednačavanje histograma, čime se izbjegava prekomjerno pojačavanje šuma i gubitak detalja.



CLAHE poboljšana slika

## **Glavne karakteristike CLAHE algoritma:**

### **1. Lokalna obrada slike**

Umjesto jedinstvenog globalnog histograma, izračunava se histogram za svaki segment slike. To omogućava bolje prikazivanje finih struktura, poput početnih promjena na listu.

### **2. Ograničavanje kontrasta (clip limit)**

CLAHE uvodi gornju granicu histogram vrijednosti kako bi se spriječilo pojačanje sume u tamnim ili svijetlim regijama. Vrijednosti iznad limita ravnomjerno se preraspoređuju.

### **3. Interpolacija između blokova**

Nakon lokalnog izjednačavanja, blokovi se spajaju uz interpolaciju kako bi se izbjegli prelazi između regija.

Matematički, clip limit  $\beta$  računa se prema formuli datoј u radu

$$\beta = \frac{M}{N} \left( 1 + \frac{\alpha}{100} (S_{max} - 1) \right),$$

gdje su M i N dimenzije regiona,  $\alpha$  factor ograničavanja, a  $S_{max}$  maksimalna vrijednost intenziteta.

Na osnovu rezultata prikazanih u radu, CLAHE značajno poboljšava vidljivost promjena na lišću i olakšava kasniju segmentaciju.

## **Klasterovanje (k-means)**

U analiziranom radu k-means ima ključnu ulogu u izdvajajući regije od interesa (ROI), odnosno oboljelog područja lista. Autori prvo transformišu sliku u Lab kolor prostoru, gdje je razdvajanje boja stabilnije, a zatim k-means klasterovanjem grupišu piksele prema sličnosti. Na taj način izdvajaju klaster koji najbolje predstavlja zahvaćeno područje, čime se uklanja uticaj pozadine i nebitnih djelova slike.

### **Proces u radu uključuje:**

- nasumičan odabir početnih centroida,
- dodjela svakog piksela najbližem centru,
- ažuriranje centroida,
- iteriranje do konvergencije,
- izdvajanje klastera koji predstavljaju zahvaćena područja.

Znači k-means ovdje ne služi za klasifikaciju bolesti, već za usmjeravanje pažnje algoritma na najrelevantniji dio slike. Ova segmentacija je presudna jer omogućava da SVM klasifikator u sledećoj fazi radi nad jasno definisanim i vizuelno dosljednim karakteristikama, što direktno doprinosi visokoj tačnosti klasifikacije prikazanoj u radu.

## **Klasifikacija (SVM)**

SVM (Support Vector Machine) predstavlja jedan od najpouzdanijih algoritama za klasifikaciju u oblastima obrade slike i prepoznavanje karakteristika. Njegova glavna prednost je sposobnost da pronađe optimalnu granicu razdvajanja između klasa, čak i u slučajevima kada je broj uzoraka mali, što ga čini pogodnim za dataset korišćen u analiziranom radu.

SVM funkcioniše tako što pronalazi hiperravan koja razdvaja klase na takav način da je rastojanje izmedju hiperravnih i najbližih uzoraka iz svake klase maksimalno. Ti najbliži uzorci nazivaju se support vectors i upravo oni određuju konačnu poziciju granice razdvajanja.

U datom naučnom radu autori koriste linearni kernel, jer su nakon CLAHE poboljšanja i k-means segmentacije dobijene karakteristike dovoljno dobro razdvojive u originalnom prostoru podataka.

Za binarnu klasifikaciju, optimalna hiperravan definiše se uslovom

$$\alpha \cdot y_i + b = \begin{cases} \geq +1, & x_i = +1 \\ < +1, & x_i = -1 \end{cases}$$

Gdje je  $y_i$  ulazni vector,  $x_i$  klasa, a  $\alpha$  i  $b$  parametri modela.

Ograničenje margine

$$\alpha_i + b \leq \pm 1$$

Kernel funkcija

$$k(x_i, x_j) = \varphi(x)\varphi(x_j)$$

Funkcija odluke

$$f(x) = sign\left(\sum_i \alpha_i y_i \varphi(x)\varphi(x_i) + bias\right)$$

## 6. Analiza rezultata

U naučnom radu primjenjena metoda (CLAHE + k-means + SVM) evaluirana je na Kaggle datasetu iz kojeg su izdvojene slike tri kategorije lišća: zdravo lišće, lišće zaraženo antraknozom i lišće zaraženo bakterijskom pjegavošću.

Autori analiziraju uspješnost klasifikacije kroz dva tipa rezultata:

1. Tačnost klasifikacije po pojedinačnim bolestima,
2. Ukupna tačnost pri različitim podjelama trening/test skupa.

### Tačnost klasifikacije po bolestima

Tabela prikazuje rezultate za dvije glavne bolesti obuhvaćene studijom

Vrsta bolesti	Tačnost (%)
Antraknoza	96.9892
Bakterijska pjegavost	97.0847

Ovi rezultati pokazuju da model ostvaruje gotovo identičnu uspješnost na obje vrste bolesti, što ukazuje da izdvajanje ROI područja putem k-means segmentacije i poboljšanje kontrasta CLAHE metodom omogućava prepoznatljive i stabilne karakteristike za klasifikaciju SVM-om.

### Ukupna tačnost pri različitim podjelama skupa podataka

Autori testiraju model na tri odnosa podjela trening/test slika:

Odnos trening/test	Tačnost (%)
70/30	99.3
60/40	97.03

Najveća tačnost (99.3%) ostvarena je kada je model treniran na većem broju slika (70%). Smanjivanjem trening skupa na 60% tačnost opada na 97.03%.

### Poređenje sa drugim metodama iz literature

Autori porede svoj pristup sa različitim tehnikama iz prethodnih radova, uključujući: GLCM, SVM, DL i ML.

Metoda	Tacnost (%)
GLCM	90
SVM	93
DL	95
ML	89
CLAHE + k-means + SVM	97.03

Iz ovih rezultata se vidi da predloženi model postiže najveću tačnost među uporednim tehnikama prikazanim u naučnom radu.

Autori objašnjavaju da konkurentne metode uglavnom ne koriste prilagodljivo poboljšanje kontrasta ili se oslanjaju na jednostavnije verzije k-means segmentacije, što dovodi do slabije detekcije oboljelih područja i niže ukupne preciznosti.

Rezultati iz naučnog rada pokazuju da kombinacija CLAHE + k-means + SVM daje:

- odličnu tačnost (97 - 99.3%),
- vrlo stabilnu klasifikaciju po klasama,
- superiorne performanse u poređenju sa drugim klasičnim metodama,
- računski efikasan model pogodan za jednostavne uređaje.

## 7. Ograničenja i moguća poboljšanja

Iako metoda predstavljena u analiziranom istraživačkom radu prikazuje visoku tačnost u identifikaciji bolesti lišća čilija, postoje određena ograničenja koja utiču na praktičnu primjenu u realnim uslovima. Najvažnije ograničenje odnosi se na korišćeni dataset, koji je sniman u kontrolisanom okruženju i stabilnim osvjetljenjem. Takvi uslovi olakšavaju segmentaciju i klasifikaciju, ali ne odražavaju kompleksnost stvarnih poljoprivrednih scena gdje su prisutni šum, sjenke, preklapanje lišća i različite pozadine. Stoga nije izvjesno da bi model zadržao jednaku tačnost pri radu sa slikama snimljenim u terenskim uslovima.

Ograničenja se javljaju i kod korišćenih metoda. CLAHE može pojačati šum u tamnim regijama slike, što otežava naknadnu segmentaciju. K-means klasterovanje je osjetljivo na početne vrijednosti centroida i ne garantuje optimalno izdvajanje oboljelih područja u slučaju složenijih slika. SVM metod pokazuje dobre rezultate na malim datasetovima, ali može izgubiti efikasnost ukoliko se poveća broj klasa ili različitost podataka.

Autori ne uključuju naprednije evaluacione metrike poput F1-score, matrice konfuzije ili ROC krivih, koje bi omogućile detaljniju analizu ponašanja modela po pojedinačnim klasama. Takođe, iako se u radu navode savremene tehnike dubokog učenja kao što su CNN modeli, ne vrši se direktno eksperimentalno poređenje sa njima, pa nije moguće jasno procjeniti prednosti i ograničenja predložene metode u odnosu na moderne pristupe.

Na osnovu navedenih ograničenja, mogu se izdvojiti pravci za buduća poboljšanja. Dataset bi trebalo proširiti slikama snimljenim u realnim uslovima, uz uključivanje različitih pozadina, osvjetljenja i faza rasta biljke. Uvodjenje naprednih dubokih arhitektura kao što je EfficientNet ili neka druga arhitektura duboke neuronske mreže omogućilo bi precizniju segmentaciju i pouzdanije prepoznavanje bolesti. Preporučljivo je i sprovesti terenska testiranja modela i razviti praktične softverske ili mobilne aplikacije, koje bi omogućile primjenu sistema u stvarnim poljoprivrednim okruženjima.

## 8. Zaključak

Bolesti lišća čilija predstavljaju značajan problem u poljoprivrednoj proizvodnji, a pravovremeno prepoznavanje ključno je za sprječavanje gubitaka u prinosu. Tradicionalne metode dijagnostike u velikoj mjeri zavise od subjektivne procjene stručnjaka, što ih čine sporim i nepouzdanim u uslovima intenzivne proizvodnje. Zbog toga se poslednjih godina sve vise pažnje posvećuje automatizovanim sistemima zasnovanim na obradi slike i mašinskom učenju.

Analizirani naučni rad pokazuje da kombinacija CLAHE metode poboljšanja kontrasta, k-means segmentacije i SVM klasifikacije može ostvariti visoku tačnost u identifikaciji bolesti lišća čilija, dostižući vrijednosti od 97% do 99.3% u zavisnosti od podjele trening i test skupa. CLAHE efikasno povećava vidljivost oboljelih područja, k-means precizno izdvaja regije od interesa, dok SVM pokazuje odlične performanse na malom dataset-u. Rezultati jasno ukazuju da pravilno kombinovane klasične metode i dalje mogu biti konkurentne savremenim dubokim modelima, posebno u okruženjima sa ograničenim resursima.

Ipak, rad ima i određena ograničenja. Dataset je mali i sniman u kontrolisanim uslovima, pa nije sigurno kako bi model reagovao na slike nastale u realnim poljoprivrednim uslovima, gdje su prisutne razlike u osvjetljenju, pozadini i stepenu oštećenja lista. Takođe, autori ne pružaju dodatne evaluacione metrike što otežava detaljnu procjenu grešaka modela.

Uprkos tome, istraživanje predstavlja značajan doprinos u oblasti automatizovane detekcije biljnih bolesti. Predloženi pristup je računski lagan, jednostavan za implementaciju i pokazuje potencijal za kasniji razvoj praktičnih aplikacija, uključujući mobilne sisteme za dijagnostiku bolesti u realnom vremenu. Budući rad treba da uključi veće i realističnije skupove podataka, testiranje u terenskim uslovima i primjenu naprednih dubokih modela koji bi dodatno poboljšali preciznost i otpornost sistema.

## **9. Prijedlog projekta**

Na osnovu analiziranog istraživačkog rada, planira se izrada projektnog modela koji koristi iste metode kao u originalnoj studiji: CLAHE za poboljšanje kontrasta, k-means klasterovanje za segmentaciju slike i SVM algoritam za klasifikaciju bolesti lišća čilija. U projektu će se koristiti isti Kaggle dataset kao i u istraživačkom radu, s tim što će za potrebe implementacije biti izdvojene različite kategorije lišća:

- zdravo lišće,
- lišće zaraženo pepelnicom,
- lišće zaraženo plamenjačom.

### **Planirane metode i njihova uloga u projektu:**

#### ***CLAHE – poboljšanje kvaliteta slike***

U projektu će se CLAHE koristiti kao prvi korak obrade slike, sa ciljem da se poboljša vidljivost oboljelih regija na listu. Ova metoda povećava lokalni kontrast i omogućava jasnije razlikovanje oboljelih područja od ostatka lista. Bolji kontrast direktno utiče na efikasnost segmentacije, jer omogućava preciznije izdvajanje regija.

#### ***K-means klasterovanje – segmentacija oboljelih regija***

K-means će se koristiti da se izdvoji regija od interesa (ROI) koja sadrži karakteristične promjene na lišću. Pored segmentacije, u projektu će biti implementirane i funkcije za vizuelni prikaz rezultata klasterovanja, kako bi se jasno vidjelo da li algoritam pravilno identificuje relevantna područja. Ovakav grafički prikaz olakšava analizu i validaciju ponašanja modela.

### ***SVM klasifikacija – prepoznavanje vrste bolesti***

Na osnovu segmentacije regije i poboljšanja slike, SVM klasifikator će biti treniran da prepoznae kojoj klasi bolest pripada. Kako se u istraživačkom radu pokazalo da SVM postiže vrlo visoku tačnost pri radu sa manjim datasetovima, ova metoda je idealna za ponavljanje kompletног sistema bez potrebe za dubokim učenjem. SVM će klasifikovati sliku u jednoj od tri kategorije: zdrav list, list zaražen pepelnicom ili plamenjačom.

Cilj projekta je provjeriti koliko se rezultati iz istraživačkog rada mogu ponavljati u praksi, kao i procjeniti koliko se izabrane kategorije mogu efikasno klasifikovati korišćenjem klasičnih metoda obrade slike i mašinskog učenja.

## Reference

- [1] Identification of chilli leaf disease using contrast limited histogram equalisation and k-means clustering.

[https://www.researchgate.net/publication/394631080\\_Identification\\_of\\_chilli\\_leaf\\_disease\\_using\\_contrast\\_limited\\_histogram\\_equalisation\\_and\\_k-means\\_clustering](https://www.researchgate.net/publication/394631080_Identification_of_chilli_leaf_disease_using_contrast_limited_histogram_equalisation_and_k-means_clustering)

- [2] Kaggle Dataset: Chili Plant Disease Dataset. [Chili Plant Disease](#)