**LICEUL TEORETIC ION LUCA**

**I.A.- Identificare microorganisme**

**SUPERVIZORI:**

**PROFESOR BARCĂU MIRELA**

**AUTORI:**

**LUPU SEBASTIAN-RADU**

**LELIUC ROBERT-CONSTANTIN**

**2021**

**ARGUMENT**

**Inteligenta artificiala reprezinta capacitatea unui sistem de a interpreta corect datele externe, de a invata din astfel de date și de a folosi ceea ce a invatat pentru a-si atinge obiective și sarcini specifice printr-o adaptare flexibilă.**

**Istoria inteligenței artificiale așa cum este ea percepută în ziua de astăzi  tehnologie care gândește ca un om începe, însă, în a doua parte a secolului trecut. Odată cu invenția computerului digital în anii ‘40, ideea construcției unei minți artificiale, electronice, a prins tot mai mult contur. Din 1949 si pana in 1960, Arthur Samuel a lucrat intens la dezvoltarea inteligenței artificiale, de la simpla recunoaștere a unor modele, la învățarea din experiența. Spre deosebire de percepția generală, inteligența artificială nu se limitează doar la IT sau industria tehnologică, ci este utilizată pe scară largă în alte domenii, cum ar fi medicină, mediul de afaceri, sistemul educațional și producție. Un sistem A.I. poate invata prin doua metode principale: Machine Learning sau Deep Learning. De exemplu, Machine Learning se axează pe construcția sistemelor care pot învăța sau își pot îmbunătăți performanțele în funcție de datele pe care le procesează, în timp ce Deep Learning, se referă așa cum spune și denumirea la o învățare mai profundă, care are la bază metode de învățare automată, bazate pe rețele neuronale**

**Am ales sa facem o inteligenta artificiala de tip Machine Learning capabilă să identifice orice fel de tardigrada.**

**Metodele utilizate de machine learning permit eliminarea necesității programării unei mașini pentru fiecare acțiune dintr-o lista de posibilități și stabilirea unor reacții pe care inteligenta mașinii ar trebui sa le aibă.**

**În prezent, metodele de machine learning sunt din ce în ce mai sofisticate, fiind integrate, de exemplu, într-o serie de aplicații medicale complexe, cum ar fi diagnosticarea depresiei pe baza unor modele de vorbire, sau identificarea persoanelor cu tendințe de sinucidere.**

**Intrucat acest tip de inteligenta artificiala este capabilă sa invete singura din o mulțime de date, cel mai important pas a fost colectarea datelor cu ajutorul cărora am reușit sa facem o inteligenta artificiala capabilă să identifice o tardigrada.** Am ales să facem acest program de recunoaștere a tardigradelor, deoarece obținerea datelor a fost mult mai facilă, acestea fiind microorganisme care trăiesc în medii de viață extreme iar mostrele le-am prelevat din jurul liceului, tardigradele putând fi găsite oriunde. Aceste mostre le-am lăsat înmuiate în apă, timp de cateva ore, după care le-am pus la microscop, putand identifica astfel tardigrade. Am făcut multe poze atat cu tardigrade, cât și fără tardigrade, pentru a învăța programul cât mai bine. Acesta trebuie antrenat pentru a învăța cât mai mult, dar totuși trebuie sa se opreasca la un anumit punct, pentru a nu crea confuzii.

**1.Bază teoretică**

**1.1 Metoda Haar**

Spre deosebire de majoritatea inteligențelor artificiale, noi am folosit o metoda matematica numită Haar care nu constă în folosirea rețelelor neuronale artificiale.

Cercetătorii de detectare a obiectelor Paul Viola și Michael Jones în 2001 au propus o metodă de detectare a obiectelor într-o imagine digitală. Este una dintre primele metode capabile să detecteze obiecte efectiv și în timp real într-o imagine. Inventate inițial pentru a detecta fețele, poate fi, de asemenea, utilizat pentru a detecta alte tipuri de obiecte, cum ar fi mașinile sau aeronavele sau în cazul nostru tardigradele.

Un pas important este învățarea programului cu date pozitive care conțin obiectul și datelor negative de preferat în volume mari și în procent de 50%.Datele folosite pentru instruirea programului trebuie într-o primă fază revizuite și clasificate.După pregătirea datelor, programul va căuta să extragă niște caracteristici comune pe care apoi le va folosi pentru identificarea obiectelor.De exemplu pentru recunoasterea facială se vor identifica diferențele intensității lumini dintre nas și ochi.După această prelucrare se identifică o mulțime de caracteristici iar apoi prin antrenamentul acestul program în mai multe generații se clasifică caracteristicile după fiecare generație după frecvența acestora.Se consideră că după fiecare generație programul se îmbunătățește.După faza de învățare clasificator, diferitele praguri pentru fiecare clasificator sunt înregistrate într-un fișier de metadate XML, care vor fi exploatate în faza de detectare.

**1.2 Librăria OpenCV**

Este o librărie open-source cu rol în detectarea vizuală. A fost scrisă inițial în limbajul C++, apoi adaptată pentru limbajele C, C#, Python și Java.Această librărie poate fi folosită pentru sisteme de recunoaștere facială, recunoașterea gesturilor, obiectelor urmărirea și estimarea mișcărilor. Au început demersurile de a dezvolta această librărie în anul 1999 ca o inițiativă a departamentului de cercetare a gigantului Intel.Prima versiune beta a fost lansată în anul 2001 iar apoi o variantă finală 2006. Numele OpenCV vine de la cuvântul în engleză “open” care face referire că această librărie este cu sursă deschisă orice poate să o folosească și de la prescurtarea “CV” care vine de la “computer vision” din engleză.OpenCV rulează pe următoarele sisteme de operare desktop: Windows, Linux, macOS, FreeBSD, NetBSD, OpenBSD iar pe partea mobilă OpenCV rulează pe următoarele sisteme de operare mobile: Android, iOS, Maemo,BlackBerry 10. Utilizatorii pot prelua versiunea oficială din SourceForge sau ultimele coduri sursă din GitHub. OpenCV utilizează CMake.[1]// asta necesită citare de la https://ro.wikipedia.org/wiki/OpenCV

**1.3 Arhitectura cascadei**

Aceasta este un pas important în recunoașterea datelor. Este o funcție recursivă ce face parte din metoda Viola Jones. Poza inițial împărțită în subcadrane, iar în fiecare stagiu al acestui funcții se verifică folosind un clasificator considerat puternic.În medie doar 0.01% din cadrane sunt pozitive.Toate cadranele au un timp de evaluare identic. În fiecare stagiu se alege un nou clasificator care are rolul de a elimina elementele fals pozitive.În a doua etapă se consideră că este o rată de identificare de 100% cu 50% de fals pozitive.

**1.4 AdaBoost**

 Este o tehnică de inteligență artificială folosită în arhitectura cascadei pentru a combina mai multe caracteristici slabe într-un clasificator puternic.Un alt element important care trebuie subliniat este acela că Viola şi Jones au utilizat AdaBoost

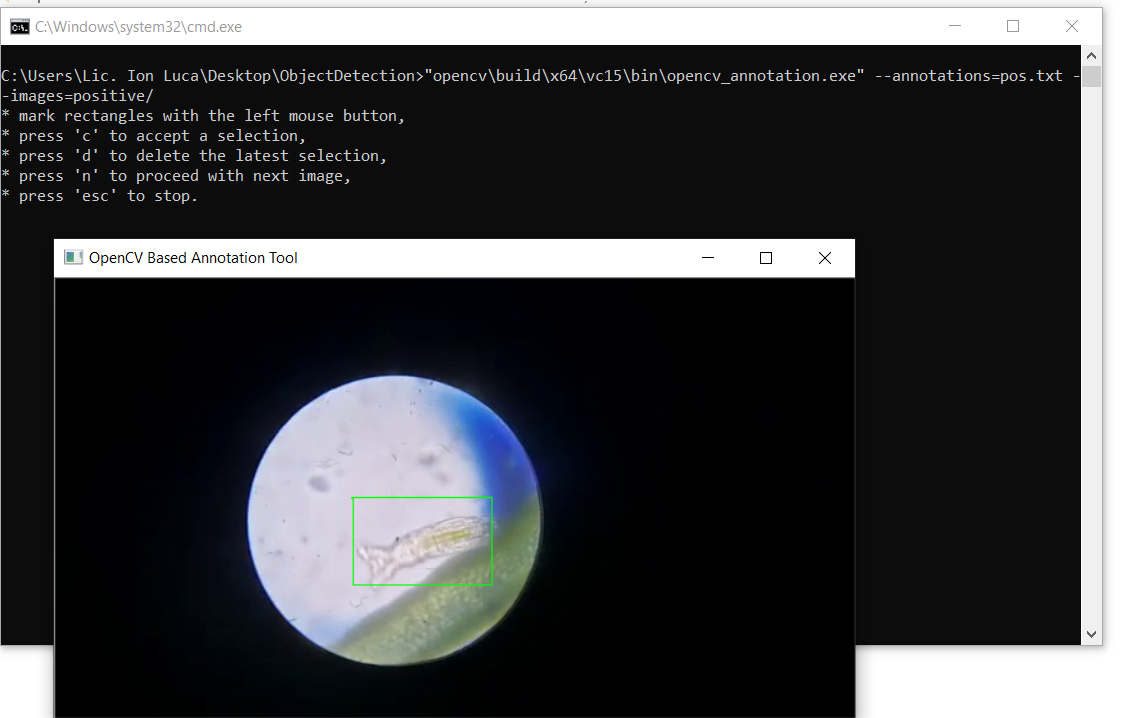
nu numai pentru antrenarea clasificatorului, dar şi pentru selectarea unui număr considerabil mai mic dintre toate caracteristicile iniţial definite, care să fie utilizate în cele din urmă.[2]//Revista Română de Informatică şi Automatică, vol. 23, nr. 2, 2013

**2.Antrenarea programului**

**2.1Procesul de prelucrare și identificare a datelor**

   După prelevarea datelor brute le-am clasificat în pozitive și negative, am folosit în jur de 600 de imagini.Am luat fiecare imagine pozitiv, analizat-o și am încadrat obiectul sau obiectele într-un dreptunghi care conține toți pixeli acestuia și am memorat numărul obiectelor și poziția acestora. Am făcut acest proces pentru a-l ajuta să identice o tardigradă şi pentru a salva coordonatele dreptunghiului în care a fost încadrat pentru a-l putea antrena mai târziu să încadreze singur o tardigradă în interiorul unui dreptunghi.

 Putem observa procesul în figura 1.



**Figura 1**

**2.2 Procesul de antrenare al programului**

Folosind un script adică un fișier cu extensia .bat am antrenat programul să identifice tardigradele folosind datele oferite acestuia adică 400 de imagini pozitive, și am ajuns la concluzia că numărul optim de generații este de 20, tot ce este peste, are o marjă de eroare mai ridicată deoarece nu am avut un ansamblu de date destul de mare, iar folosind un număr mai scăzut de generații, programul nu are timp să rafineze toate datele.

**3. Aplicații**

**3.1Prezentarea programului**

Programul este alcatuit din 4 subprograme si 3 scripturi. Doua dintre subprograme sunt folosite pentru a crea date din pozele salvate, iar mai apoi pentru a le interpreta, iar celelalte 2 subprograme sunt folosite pentru a crea date prin intermediul camerei, iar mai apoi de a le interpreta tot prin intermediul camerei. Cele 3 scripturi sunt folosite pentru a antrena programul, adică în exemplul nostru pentru a-l invata sa identifice o tardigrada.

pui poză cu folderu

**3.2 Subprograme**

**3.2.1 Subprogramul de prelucrare a datelor din galerie**

Prima oară apasand tasta q programul va șterge toate pozele din folderele positive și negative.Apoi subprogramul va prelua toate pozele din galerie, după care le va împărți în pozitive și negative. Astfel dacă identificăm într-o poza o tardigrada, vom apăsa tasta p pentru a o salva în folderul positive, altfel vom apăsa tasta n pentru a o salva în folderul negative. Cele negative reprezinta poze fara tardigrade si nu necesita niciun fel de prelucrare. Cele pozitive, în schimb necesita prelucrare în plus și anume faptul ca tardigradele trebuie încadrate manual într-un dreptunghi, așa cum se vede și în figura 1. Coordonatele dreptunghiului vor fi salvate într-un fișier de tip .txt și vor fi folosite mai târziu

**3.2.2 Subprogramul de identificare a datelor din galerie**

**3.2.3 Subprogramul de prelucrare a datelor cu date din camera foto**

**3.2.4 Subprogramul de identificare a datelor în direct**

**3.3 Scripturi**

**3.3.1 Adnotarea pozitivelor**

     Acest script este esențial în procesul de prelucrare a datelor pozitive. Cu acest program am putut pune un dreptunghi în jurul tardigradei sau obiectului dorit să fie identificat și să stocăm informația aceasta și să transferăm datele pozitive în folderul cu imagini pozitive.Putem vedea procesul de punere în dreptunghiuri a tardigradelor în figura 1.

**3.3.2 Atribuirea pozitivelor în fișierul cu extensia .xml**

    Având în vedere că programul nu poate funcționa fără fișierul cu extensia .xml, am fost nevoiți să preluăm datele pozitive din folderul positives și apoi să creăm acest fișier de tip .xml am creat acest script pentru a nu face acest lucru manual deoarece este un proces îndelungat .

**3.3.3 Antrenarea programului**

   Am scris acest script deoarece dacă programul nu este antrenat cu datele positive acesta nu este în stare să identifice tardigradele sau obiectele dorite.Acest script după rulare, trebuie să îi fie dați 2 parametri, numărul de generați dorite și numărul de poze pozitive.

**3.4 Plasticitatea programului**

 Cel mai remarcabil lucru al acestui program este capacitatea de adaptare la datele acordate.Dacă dorim să învățăm acest program să identifice obiecte precum masini , avioane sau mai multe microorganisme cum am dorit în prima fază a acestui proiect însă timpul nu ne-a lăsat această posibilitate, trebuie doar să prelucrăm poze cu lucrul dorit și să parcurgem aceeași etape iar programul va putea recunoaște tardigradele și celelalte obiecte.