

# VEDEREA ARTIFICIALĂ UTILIZATĂ PENTRU PERSOANELE CU ALZHEIMER

Cătălina Bejan<sup>1\*</sup>, Dan Munteanu<sup>2\*</sup>

\*Universitatea „Dunărea de Jos” din Galați,  
Departamentul de Calculatoare și Tehnologia Informației, România

<sup>1</sup>e-mail: catalina\_bejan@yahoo.com

<sup>2</sup>e-mail: dan.munteanu@ugal.ro

## ABSTRACT

*Persoanele care suferă de Alzheimer sunt predispuse riscului de subalimentație, supraalimentație și deshidratare deoarece pierderea memoriei de scurtă durată poate crea confuzii. Aceștia au nevoie de un îngrijitor care să se asigure că respectă mesele principale ale zilei și că se hidratează în mod corect. Scopul acestei lucrări este prezentarea necesității sistemului de supraveghere dezvoltat care are rolul de a le recăpăta independența prin intermediul inteligenței artificiale. Sistemul este bazat pe un concept inedit care implică utilizarea inteligenței artificiale pentru recunoașterea activității umane din video. Sistemul identifică momentele în care persoana supravegheată se hrănește sau se hidratează și îi amintește prin mesaje sonore că a uitat să mănânce, să bea sau mănâncă prea mult. Totodată permite supravegherea și gestionarea programului de nutriție de la distanță din partea unui îngrijitor. Lucrarea cuprinde studiul, căutarea, antrenarea și utilizarea modelelor și algoritmilor specifici domeniului de vedere artificială în scopul clasificării imaginilor, detecției obiectelor în imagini și recunoașterii activității umane în video. Studiul realizat denotă faptul că inteligența artificială utilizată în scopul ajutorării persoanelor bolnave de Alzheimer este un concept nou, dar de real ajutor atât pentru ei, cât și pentru îngrijitorii lor.*

**CUVINTE CHEIE:** Inteligență Artificială, Alzheimer, Vedere computerizată, Sistem de supraveghere, Clasificarea imaginilor, Detecția obiectelor în imagini, Recunoașterea activității umane în video

## 1. INTRODUCERE

Așa cum corpul uman se schimbă odată cu vârsta, la fel reacționează și creierul. Vârsta înaintată este cel mai mare factor de risc al apariției bolii Alzheimer, majoritatea cazurilor dezvoltându-se după vârsta de 65 de ani. Totuși s-au înregistrat cazuri incipiente ale bolii și sub această vârstă demonstrând că nu afectează doar persoanele vârstnice, dar nici nu este o etapă firească a îmbătrânirii. [8]

Demența este un termen general folosit pentru definirea unui set de anomalii care apar la instalarea bolii precum afectarea memoriei și a capacităților de gândire. Boala Alzheimer se înregistrează în 60-80% din cazurile de demență, această afecțiune provocând pierderea treptată a abilității de a funcționa independent. [8]

Sunt 50 de milioane de persoane care suferă de demență în întreaga lume și se estimează ca acest număr să se dubleze la fiecare 2 decenii, ajungând la 152 de milioane până în 2050. [9] Această creștere alarmantă a numărului de bolnavi a provocat în rândul oamenilor de știință o cursă contra cronometru pentru

găsirea unor tratamente eficiente de prevenire și ameliorare a simptomelor pe care boala le presupune.

**Error! Reference source not found.**

Fiind o boală progresivă, în faza de început boala se manifestă prin pierderi ușoare de memorie, dar se agravează în timp până la nivelul la care persoana afectată nu mai poate comunica prin cuvinte sau răspunde la stimuli din mediu. [8]

Boala Alzheimer poate cauza tulburare de alimentație, pacienții ajungând fie să uite să mănânce, provocând subnutriție, fie să mănânce de prea multe ori pe zi, provocând supraalimentație. **Error! Reference source not found.**

Pentru persoanele cu Alzheimer, riscul de deshidratare sau alimentație nesănătoasă este mult mai mare, acestea uitând chiar să mănânce sau să bea apă, chiar dacă s-a instalat senzația de sete, sau foame. Totodată, bolnavii sunt predispuși să mănânce mai mult decât este necesar sau mai des din cauza senzației de confuzie pe care o resimt, deoarece nu se pot baza pe memoria de scurtă durată. **Error! Reference source not found.**

Din aceste cauze, posibilitatea de a trăi independent devine aproape imposibilă, iar nevoia unui îngrijitor este esențială.

Astfel, apare nevoia unui sistem care să monitorizeze activitățile bolnavului pentru a urmări momentele în care mănâncă, bea și a-l atenționa atât pe el cât și pe îngrijitorul său când este important ca bolnavul să bea apă, să mănânce sau să se oprească din mâncat în scopul evitării deshidratării, subnutriției sau supranutriției.

#### A. Sisteme existente

Sistemele existente, concepute pentru a veni în ajutorul persoanelor care suferă de demență și Alzheimer, constau în sisteme de localizare a persoanei, sisteme de aducere aminte a unor activități precum luarea medicamentelor, sisteme de supraveghere video pentru urmărirea activităților bolnavului de către îngrijitor.

- Aplicații de tip memento:

„Pill Reminder” (iOS) este o aplicație de tip memento care alertează bolnavul că trebuie să își ia medicamentele la o oră fixă. Aceste alarme sunt setate de către un îngrijitor, iar aplicația permite logarea manuală făcută de către bolnav a momentului când pastila a fost administrată.

„It's Done!” (iOS) este o aplicație de tip memento în care îngrijitorul poate seta o serie de activități ce trebuie îndeplinite într-o zi, iar bolnavul le poate bifa ca fiind satisfăcute după ce le efectuează.[10]

- Sisteme de supraveghere:

Sistemele de supraveghere existente permit transmiterea fluxului video în timp real prin internet către îngrijitor, însă pentru a fi eficiente este nevoie ca îngrijitorul să monitorizeze constant activitatea bolnavului prin intermediul acestora, adică să privească fluxul video în continuu. Camerele de supraveghere „Foscam” permit acest lucru, dar și posibilitatea comunicării prin sunete între bolnav și îngrijitor. Camerele „Y-cam” permit supravegherea declanșată de mișcare. Alte camere de supraveghere în timp real sunt „Netgear Arlo” și „Baby cam”. **Error! Reference source not found.**

Din sistemele și aplicațiile menite pentru a ajuta persoanele bolnave de Alzheimer se poate observa că deocamdată nu există nicio aplicație sau sistem care să monitorizeze activitățile bolnavului în vederea prevenirii deshidratării, subalimentării sau supraalimentării. Totodată, nu există o aplicație care să permită îngrijitorilor să renunțe la monitorizarea constantă a activităților bolnavilor de care au grijă deoarece nu există un sistem cu recunoaștere automată prin inteligența artificială a activităților conceput pentru acestia.

#### B. Servicii

Sistemul de supraveghere dezvoltat permite monitorizarea și recunoașterea activității umane, transmiterea în timp real a înregistrării, salvarea momentelor în care persoana supravegheată mănâncă

sau bea într-o bază de date și notificarea bolnavului și a îngrijitorului atunci când bolnavul ar trebui să mănânce, să bea sau să se oprească din mâncat.

Astfel, persoana bolnavă de Alzheimer își poate recăpăta independența prin faptul că nu este necesar să trăiască în aceeași casă cu îngrijitorii săi.

Totodată, îngrijitorul se bucura de o siguranță în ceea ce privește nevoile și sănătatea bolnavului prin faptul că nu mai este necesar să îi monitorizeze constant activitatea, de acest lucru se va ocupa sistemul de supraveghere bazat pe vedere artificială. De asemenea, în cazul în care îngrijitorul dorește să monitorizeze activitățile bolnavului, o poate face de la distanță prin simpla conexiune la internet de pe orice dispozitiv care are acces la un browser de internet.

#### C. Tipuri de utilizatori ai aplicației

- Persoana bolnavă de Alzheimer
- Persoana care are grijă de un bolnav de Alzheimer

Lucrarea cuprinde studiul, căutarea, antrenarea și utilizarea modelelor și algoritmilor specifici domeniului de vedere artificială în scopul clasificării imaginilor, detecției obiectelor în imagini și recunoașterii activității umane în video.

## 2. IMPLEMENTAREA SISTEMULUI

Inteligența artificială a ajuns la nivelul actual de dezvoltare datorită dorinței continue a experților din domeniu de a construi mașinării care să poată realiza tot ceea ce o ființă umană este capabilă să facă.

Acest domeniu este în continuă evoluție, fiind deja un ajutor în viața de zi cu zi a oamenilor (prin asistenți personali pe smartphone precum Siri, Alexa, Google Assistant) și chiar depășind capacitățile fizice umane în unele situații (exploatarea substanțelor chimice dăunătoare, lucrul în fabricile mari de producție). [1]

Inteligența artificială și aplicabilitatea ei în activitățile cotidiene au progresat în special în ultimul deceniu. Scopul dezvoltării inteligenței artificiale constă în crearea unor mașinării care să fie apte să perceapă, înțeleagă, interpreteze, determine, raționalizeze, recunoască și să reacționeze la stimulii din mediu, la fel ca o ființă umană. [1]

#### A. Inteligența artificială modernă

Descoperirile din 2012 (printre care și o rețea neuronală construită de cercetătorul Geoffrey Hinton și echipa lui cu care au câștigat un concurs internațional de vedere artificială)[3] au determinat progresele de astăzi cu privire la vederea artificială, recunoașterea limbajului și mașinile cu conducere autonomă. Acești factori au schimbat felul în care inteligența artificială este văzută în ziua de astăzi.[1]

Subdomeniile inteligenței artificiale cuprind:

- prelucrarea limbajului natural (se ocupă cu analiza și înțelegerea informațiilor din limbajul natural din documente)

- robotică (se ocupă cu proiectarea, programarea și fabricarea roboților)
- învățarea automată (capacitatea unei mașini de a învăța) și învățarea profundă (subdomeniu al învățării automate)
- sistemele expert (rezolvă probleme complexe în urma analizării unei baze de cunoștințe care conține reguli și generării unei concluzii)
- recunoașterea vocală (se ocupă cu recunoașterea cuvintelor vorbite și a vocii)
- automatizarea proceselor (se ocupă de automatizarea roboților prin combinație cu inteligența artificială)
- vederea artificială (abilitatea mașinilor de a vedea și înțelege, interpreta ceea ce văd).

Pentru construirea unui proiect bazat pe inteligență artificială este adesea necesară folosirea mai multor subdomenii, ele fiind interconectate. [1]

### B. Vedere artificială

Vederea artificială (în engleză *Computer Vision*) este un domeniu al informaticii și ingineriei care a câștigat popularitate și a obținut progrese considerabile în ultimul deceniu.

Scopul acestui domeniu este extragerea informațiilor din imagini sau clipuri video, necesare pentru a deduce o informație despre lume. [4]

Construirea unor mașini care să poată vedea, înțelege și interpreta ceea ce văd este o problemă pe care experții din domeniul inteligenței artificiale încearcă să o rezolve de aproximativ șase decenii. Ceea ce s-a dovedit a fi o problemă complexă, era considerată inițial suficient de ușor de rezolvat încât să fie desemnată ca temă de vară pentru un student. [2]

### C. Rețele neuronale artificiale

Cercetătorii inteligenței artificiale au abordat două metode pentru imitarea gândirii și deducerii rezultatului:

- metoda bazată pe reguli și
- metoda care utilizează rețele neuronale.

Metoda bazată pe reguli (sisteme expert) presupune utilizarea unei serii de condiții logice: dacă în imagine se identifică un cerc, atunci există posibilitatea ca în imagine să fie o farfurie. Dacă cercul include o altă formă în interiorul său, atunci există posibilitatea ca în imagine să fie o farfurie cu mâncare.

Metoda care utilizează rețele neuronale a fost dezvoltată după modelul biologic al creierului uman. Această abordare presupune imitarea rețelelor de neuroni care se găsesc în creierul oamenilor și al animalelor, fiind singura formă de inteligență pe care o cunoaștem. Aceste rețele neuronale au fost denumite rețele neuronale artificiale și sunt formate din straturi de neuroni artificiali, care transmit informații între ei la fel cum sunt transmise informațiile prin sinapse în mod biologic. Rețelele neuronale artificiale sunt capabile să învețe din experiență. Acest lucru presupune că dacă o rețea neuronală artificială a primit anterior mai multe exemple de imagini care conțin farfurii cu mâncare,

atunci este capabilă să identifice tipare din care să deducă dacă o imagine nemeaivăzută până atunci se aseamănă cu cele etichetate ca fiind farfurii cu mâncare. [3]

Rețelele neuronale artificiale necesită putere computațională mare, dar și seturi de date mari din care să învețe. [3] Seturile de date trebuie să fie cât mai vaste pentru a oferi perspective reale ale obiectului cu ajutorul cărora să poată fi identificat în lumea tridimensională în care trăim din orice unghi și la orice distanță este surprins în imaginea digitală (sau fisier video).

### D. Seturile de date

Odată cu avansarea tehnologiei, puterea computațională a devenit suficientă pentru folosirea rețelelor neuronale artificiale, iar seturile de date au devenit din ce în ce mai mari, multitudinea de exemple fiind necesară pentru alcătuirea experienței din care programul învață. În cazul seturilor de date alcătuite din imagini sau video, creșterea cantității de date a fost posibilă datorită faptului că majoritatea oamenilor dețin un dispozitiv cu ajutorul căruia pot fotografia și înregistra clipuri video (ex: smartphone, aparat foto-video, camera web). Adunarea respectivelor date digitale este posibilă datorită accesibilității la internet și la rețelele sociale, în cadrul cărora pot fi încărcate diferite fotografii sau videoclipuri.

### E. Apariția învățării profunde

În ciuda faptului că puterea computațională și dimensiunea seturilor de date nu mai reprezintă un impediment, rețelele neuronale artificiale necesitau multe straturi de neuroni pentru a putea fi utilizabile într-o aplicație de necesitate reală, dar acestea nu puteau fi antrenate într-un mod eficient pe măsura ce erau adăugate la rețea. Acest lucru s-a schimbat odată cu apariția învățării profunde prin descoperirile făcute de cercetătorul Geoffrey Hinton. Acesta a reușit să antreneze în mod eficient straturile de neuroni, aducând progrese considerabile pentru vederea artificială, dar și pentru recunoașterea vocală. [3]

Învățarea profundă a obținut popularitatea meritată în 2012, atunci când Geoffrey Hinton și echipa lui au câștigat concursul internațional „ImageNet Large Scale Visual Recognition Challenge” cu o rețea neuronală convoluțională construită de ei. [7]

### F. Rețele neuronale convoluționale

Rețelele neuronale convoluționale sunt dezvoltate pentru procesarea imaginilor sau a videoclipurilor și datelor similare 2D/3D. Ceea ce le diferențiază de rețelele neuronale simple este faptul că sunt alcătuite din straturi convoluționale și utilizează straturi de *pooling* pentru compresia imaginii. Pentru returnarea rezultatului, aceste rețele folosesc straturi complet conectate (toți neuronii sunt conectați). [5]

În Figura 1 este prezentată o rețea neuronală convoluțională alcătuită din două straturi convoluționale care returnează câte o imagine pentru fiecare filtru, două straturi de pooling și un strat

complet conectat pentru ieșire care clasifică în șase categorii (șase neuroni pe stratul de ieșire).

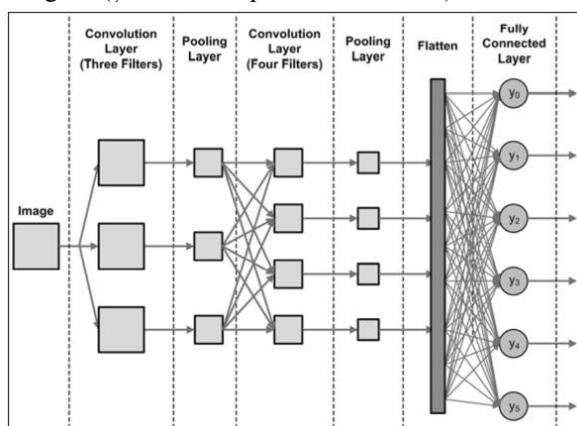


Figura 1. Rețea neuronală convoluțională [5]

Aplicații de interes în cadrul domeniului de vedere artificială:

- Clasificarea imaginilor – presupune analiza întregii imagini și clasificarea ei într-o singură categorie
- Detecția obiectelor în imagini – presupune identificarea și localizarea unuia sau mai multe obiecte, încadrându-le în chenare în cadrul unei imagini
- Segmentarea imaginilor – presupune identificarea și localizarea pixelilor care alcătuiesc obiectele dintr-o imagine
- Recunoașterea activității în video – presupune analiza unui întreg video și clasificarea lui într-o singură categorie
- Localizarea activității în video – presupune identificarea și localizarea uneia sau mai multor acțiuni dintr-un video

#### G. Clasificarea imaginilor

Pentru clasificarea imaginilor s-a efectuat antrenarea atât de la zero a rețelelor neuronale convoluționale, cât și folosirea tehnicii de învățare prin transfer de cunoștințe (în engleză *transfer learning*).

S-a efectuat modelarea și construirea unei rețele neuronale convoluționale care utilizează arhitectura modelului VGG-16 pentru clasificarea imaginilor și care este instruită anterior pe setul de date ImageNet.

Alegerea acestei arhitecturi și efectuarea clasificării binare pentru identificarea mâncării (clasa „food”) a fost efectuată în urma antrenării mai multor modele utilizând atât o arhitectură de rețea construită de la zero, cât și transfer de cunoștințe pe modele precum: MobileNet, InceptionV3 și VGG16, dar și seturi de date diferite și clase diferite pentru identificarea imaginilor de interes (cele care conțin mâncare).

#### Setul de date

Seturile de date folosite pentru antrenarea modelelor pentru clasificarea imaginilor sunt Food-101 **Error! Reference source not found.**, Furniture Detector **Error! Reference source not found.** și The

IMDB-WIKI dataset **Error! Reference source not found. Error! Reference source not found. Error! Reference source not found.**

Datorită faptului că se dorește clasificarea imaginilor ca fiind cu mâncare sau fără mâncare, seturile de date menționate anterior cuprind fie imagini cu tipuri de mâncare, fie imagini fără mâncare întâlnite în mod frecvent (obiecte de mobilier, oameni).

#### Modelare

Pentru modelare s-a utilizat mediul Jupyter Notebook prin intermediul Colaboratory, în cadrul caruia codul și comentariile sunt structurate în celule de cod ce pot fi rulate individual sau celule de text. **Error! Reference source not found.**

Limbajul folosit pentru antrenare este Python, iar dependențele de biblioteci software folosite sunt NumPy, care este utilizat pentru operațiile numerice (inclusiv funcții matematice pentru lucrul cu liste și matrici multidimensionale), Keras pentru modelarea rețelei neuronale artificiale, Glob pentru obținerea căilor de fișiere din directorul specificat (obținerea listei căilor tuturor imaginilor din setul de date), Matplotlib pentru vizualizarea graficelor și Scikit-learn (cunoscută și ca *sklearn*) pentru calcularea acurateții modelului pe subsetul de testare și crearea matricei confuziilor.

În Figura 2 este prezentată împărțirea setului de date, iar fiecare clasă din subsetul de antrenare conține câte 1000 de imagini, iar fiecare clasă din subsetul de testare conține câte 250 de imagini, adică setul de date

este împărțit în 80% date pentru antrenare și 20% date pentru testare (Figura 3).

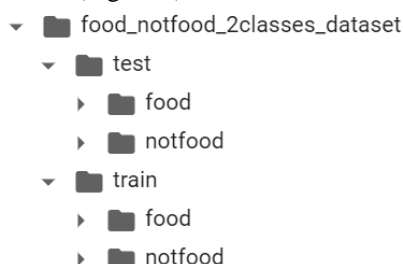


Figura 2. Structura setului de date pentru clasificare binară

Number of images per class:		
class	train	test
food:	1000	250
notfood:	1000	250
Total:	2000	500

Figura 3. Setul de date pentru clasificare binară împărțit în 80% date pentru instruire și 20% date pentru testare

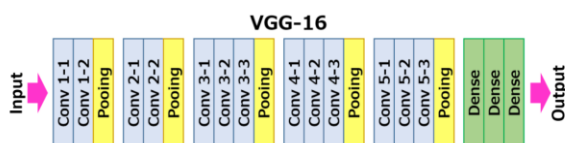


Figura 4. Blocurile convoluționale din arhitectura VGG16 [37]

### Evaluarea performanței

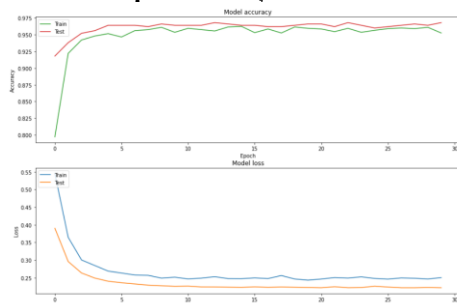


Figura 5. Graficele procesului de antrenare

În Figura 5 se observă cum acuratețea modelului (antrenat prin transfer de cunoștințe de la un model preantrenat cu arhitectura prezentată în Figura 4) pe setul de antrenare și testare a crescut în timpul antrenării, iar pierderea (memorizarea) a scăzut pe perioada antrenării, rezultând o antrenare a modelului efectuată cu succes.

Scorul de acuratețe în urma antrenării modelului este de **96,8%**.

### Predicții



Figura 6. Vizualizarea imaginilor, a predicțiilor și a etichetei corecte

În Figura 6 sunt afișate șase imagini alese aleator împreună cu eticheta lor corectă, predicția făcută de către model și scorul fiecăreia. Din cele șase imagini toate sunt clasificate corect, predicțiile fiind corespunzătoare cu eticheta lor reală.

### Matricea confuziilor

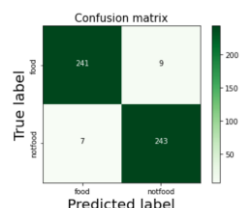


Figura 7. Matricea confuziilor

Din matricea confuziilor de la Figura 7 se observă că din 500 de imagini testate, valorile rezultate sunt în număr de:

- 241 adevărat-pozitive (modelul a prezis corect clasa pozitivă - cea căutată=„food”)
- 243 adevărat-negative (modelul a prezis corect clasa negativă – care nu e food)
- 7 fals-pozitive (modelul a prezis incorect clasa pozitivă)
- 9 fals-negative (modelul a prezis incorect clasa negativă)

În urma verificării matricei confuziilor se observă că numărul predicțiilor false este unul scăzut comparativ cu numărul total de predicții făcute, ceea ce indică faptul că modelul antrenat este unul cu o performanță ridicată.

### Concluzii asupra clasificării imaginilor

În urma efectuării verificărilor pe scorul de performanță, matricea confuziilor, vizualizarea imaginilor clasificate în mod eronat și predicția pe o imagine care nu se găsește în setul de date se pot concluziona anumite aspecte.

Modelul de clasificare binară a cărui arhitectură are la bază arhitectura modelului VGG16 și care are



transfer de cunoștințe de la același model este unul cu o acuratețe ridicată și rata de eroare scăzută, aspecte care îl transformă într-un model potrivit pentru utilizarea în viața reală.

#### H. Detecția obiectelor în imagini

Detecția obiectelor în imagini presupune recunoașterea și localizarea obiectelor în cadrul imaginii. Obiectele de interes identificate vor fi încadrate într-un chenar desenat pe imaginea pe care s-a efectuat procesul de detecție a obiectelor. (Figura 8)

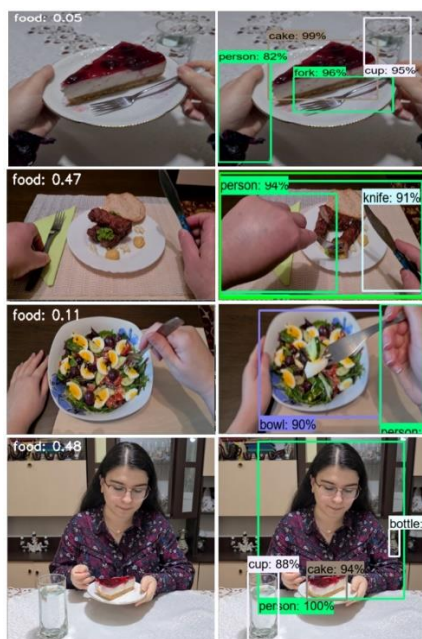


Figura 8. Diferența dintre clasificarea imaginilor și detecția obiectelor

Pentru detecția obiectelor în imagini s-a efectuat antrenarea unei rețele neuronale convoluționale prin tehnica de transfer de cunoștințe (în engleză *transfer learning*) având la bază arhitectura și cunoștințele modelului *SSD ResNet50 V1 FPN 640x640 (RetinaNet50)*.

Totodată s-a utilizat un model preantrenat care a obținut un scor de performanță ridicat pe setul de date *Coco2017*, acesta fiind modelul *Faster R-CNN ResNet50 V1 640x640*.

S-a efectuat și adnotarea unui set de date propriu, alcătuit din imagini care conțin obiecte de interes, imaginile fiind căutate și selectate manual, apoi adnotate utilizând instrumentul *LabelImg*. **Error! Reference source not found.**

#### Setul de date

Seturile de date folosite pentru antrenarea modelelor pentru detecția obiectelor sunt „*Fruit Images For Object Detection*” **Error! Reference source not found.** și „*COCO 2017 Dataset*”. **Error! Reference source not found.** În Figura 9 și Figura 10 este prezentată tehnica de adnotare a imaginilor și etichetele folosite pentru adnotare.

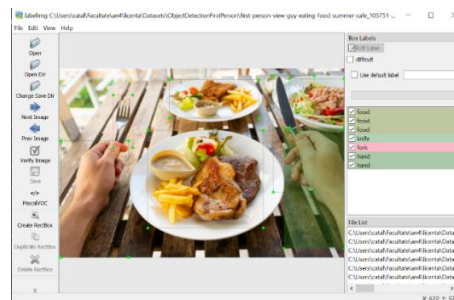


Figura 9. Adnotarea imaginilor [19]

#### Modelare

```
1 item {
2   id: 1
3   name: 'apple'
4 }
5 item {
6   id: 2
7   name: 'orange'
8 }
9 item {
10  id: 3
11  name: 'banana'
12 }
```

Figura 10. Harta etichetelor claselor de obiecte

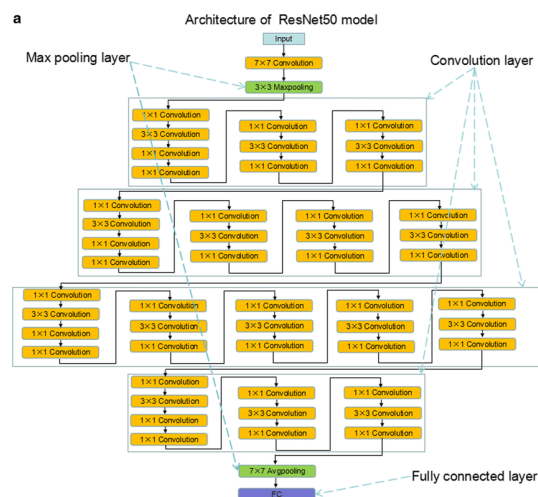


Figura 11. Arhitectura modelului ResNet50 [18]

#### Evaluarea rezultatelor

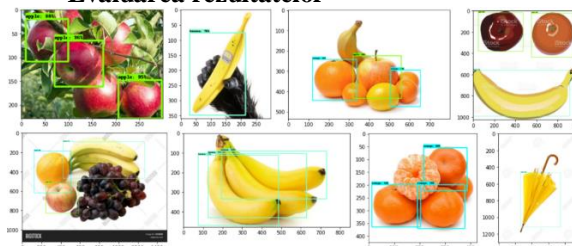


Figura 12. Rezultatul detecției obiectelor în imagini [21]

În urma testării modelului (cu arhitectura din Figura 11), modelul are o acuratețe de aproximativ 74% pe imaginile testate. În ultima imagine din Figura 12 se poate observa cum modelul a făcut o detecție eronată, considerând umbrela ca fiind o banană cu o precizie de peste 40%.

Această eroare apare din cauza setului de date pe care a fost antrenat modelul deoarece acesta conține

doar trei tipuri de clase (măr, portocală, banană), iar modelul încearcă să clasifice toate obiectele în una din categoriile pe care le cunoaște din experiență. Existența unei alte clase numită eventual „non fruit” ar fi ajutat modelul să facă o clasificare a obiectelor mai precisă.

#### Modelul preantrenat Faster R-CNN

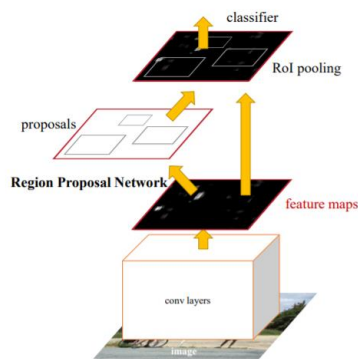


Figura 13. Arhitectura Faster R-CNN [32]

În Figura 14 sunt prezentate detecțiile obiectelor făcute cu modelul preantrenat *Faster R-CNN ResNet50 V1 640x640*, prezentat în Figura 13.

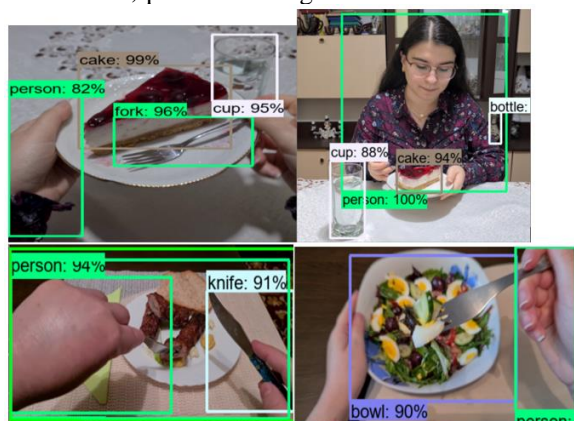


Figura 14. Detecția obiectelor în imagini

#### Concluzii asupra detecției obiectelor în imagini:

Detecția obiectelor în imagini este o metodă de mare interes pentru cerințele vremurilor în materie de vedere artificială, fiind o problemă complexă. În urma antrenării unui model prin transfer de cunoștințe și a verificării unui model preantrenat se poate concluziona că modelele pentru detecția obiectelor au o performanță ridicată, obiectele fiind detectate aproape în timp real.

#### I. Recunoașterea activității umane în video

Recunoașterea activității (evenimentelor) este o ramură a viziunii computerizate și presupune recunoașterea acțiunilor sau evenimentelor dintr-un video în funcție de context. [6]

##### Setul de date

Pentru recunoașterea activității în video, modelul a fost antrenat pe setul de date „Kinetics 400”, care, după cum sugerează și numele, conține 400 de activități umane cu cel puțin 400 de videoclipuri

fiecare, înregistrate în videoclipuri de pe YouTube. **Error! Reference source not found.**

Dintre toate cele 400 de activități, pentru această lucrare se vor lua în considerare doar activitățile legate de acțiunea de a mânca sau a bea, precum „eating burger”, „drinking”.

#### Modelare

Problema clasificării acțiunilor umane în video este o problemă mult mai complexă deoarece problema presupune analiza unui video care este alcătuit atât dintr-o componentă spațială (cadre), dar și dintr-o componentă temporală (ordinea cadrelor). **Error! Reference source not found.**

##### Primul model

Modelul are arhitectura rețelei ResNet34, care este alcătuită după cum îi spune numele din 34 de straturi convoluționale, vizibile în Figura 15.

Layer name	Resnet-34
conv1	7x7, 64, stride 2
pool1	3x3, max pool, stride 2
conv2_x	$3 \times 3, 64$ $3 \times 3, 64$ × 3
conv3_x	$3 \times 3, 128$ $3 \times 3, 128$ × 4
conv4_x	$3 \times 3, 256$ $3 \times 3, 256$ × 6
conv5_x	$3 \times 3, 512$ $3 \times 3, 512$ × 3
fc1	4x1, 512, stride 1
pool time	1x10, avg pool, stride 1
fc2	1x1, 50

Figura 15. Arhitectura ResNet34 [29]

Modelul a fost realizat odată cu lucrarea „Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?” publicată în 2018 care introduce arhitectura din Figura 16. Modelul a obținut o acuratețe de 78,4% pe setul de testare al setului de date Kinetics400. **Error! Reference source not found.**

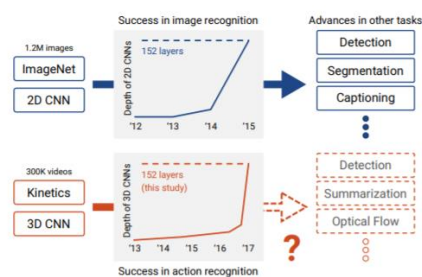


Figura 16. Comparație făcută în studiul antrenării modelului de clasificare a acțiunilor umane în video [23]

Activitatea recunoscută în Figura 17 este „tasting food”, iar această activitate este interpretată în aplicație ca o activitate generală „eat”.

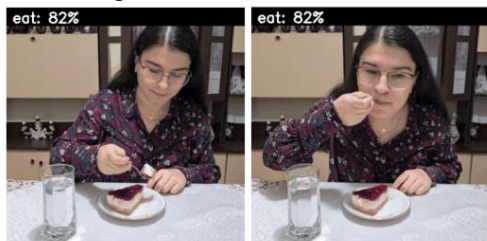


Figura 17. Cadre dintr-un video după recunoașterea activității umane

#### Al doilea model

Un alt model preantrenat testat pentru această aplicație este cel dezvoltat odată cu lucrarea „Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset”. Lucrarea a fost publicată în 2017 și a introdus o nouă arhitectură Two-Stream Inflated 3D ConvNet bazată pe inflația rețelelor convoluționale 2D, în care nucleul 2D al rețelelor profunde pentru clasificarea imaginilor a fost extins la unul 3D, făcând posibilă procesarea componentei spațiale și temporale din video. **Error! Reference source not found.** **Error! Reference source not found.**

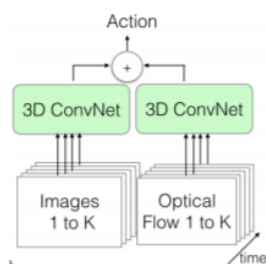


Figura 18. Arhitectura Two-Stream I3D [13]

Acest model (Figura 18) a obținut o acuratețe de 74,2% pe setul de testare al Kinetics 400. **Error! Reference source not found.**

Comparând cele două modele de recunoaștere a activității umane în video s-a ales primul model, bazat pe rețele neuronale convoluționale 3D spațiotemporale deoarece înregistrează atât o acuratețe mai mare pe setul de testare Kinetics 400, cât și o viteză de calcul mai mare, fiind mai rapid decât al doilea model.

#### Concluzii asupra recunoașterii activității umane în video:

Recunoașterea activității umane în video este una dintre cele mai complexe probleme din domeniul vederii artificiale. Aceasta presupune atât procesarea imaginilor, dar și componenta temporală dintr-un videoclip (ordinea imaginilor). Modelele acestea necesită o putere computațională mare pentru antrenare, dar și pentru testare dacă se dorește o viteză de răspuns aproape în timp real.

#### J. Echipamente hardware

Pentru realizarea dispozitivului de supraveghere video și notificare audio s-au folosit echipamentele din Figurile 19,20,21,22:

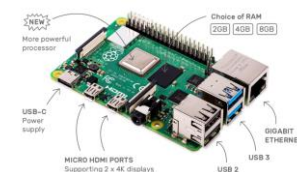


Figura 19. Raspberry Pi 4 Model B [35]

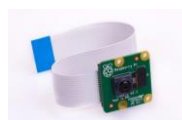


Figura 20. Raspberry Pi Camera V2 [35]



Figura 21. Alimentator Raspberry Pi 4 [35]



Figura 22. ASUS ZenPower 10050mAh [11]

Bateria externă (Figura 23) este folosită în cazul în care se dorește ca dispozitivul să fie portabil.



Figura 23. Boxa Philips BT25G/00 [12]

#### K. Aplicație de monitorizare

Pentru realizarea aplicației s-a dezvoltat o aplicație web realizată în limbajul Java, folosind servicii REST, Spring Boot, Spring Data JPA, Apache Maven, Thymeleaf, HTML, MySQL, Flask, WebSockets, ImageZMQ, OpenCV, Google Text to Speech, OMXPlayer

### 3. REZULTATE

#### Desktop

Dispozitivul are o poziție fixă (ex. pe perete) în Figura 24 și Figura 25.

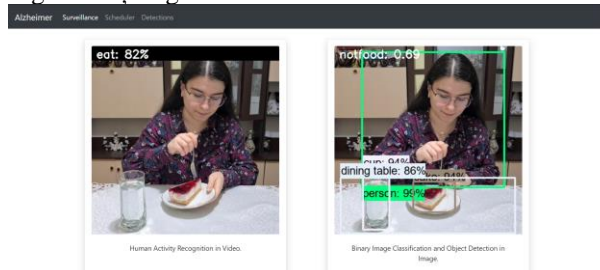


Figura 24. Surveillance 1



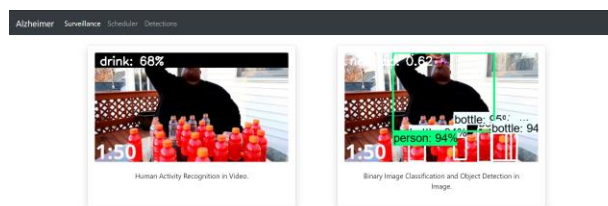


Figura 25. Surveillance 2

Dispozitivul este mobil și filmează din perspectiva în care persoana supravegheată privește deoarece îl poartă în permanență ca în Figura 26 și Figura 27.

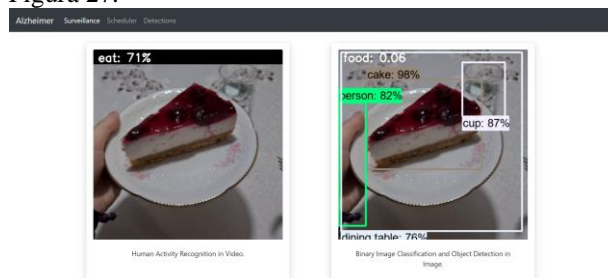


Figura 26. Surveillance 2

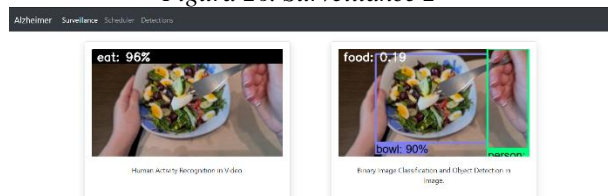


Figura 27. Surveillance 3

Aplicația poate fi vizualizată în formă mobilă ca în Figura 28, Figura 29 și Figura 30.

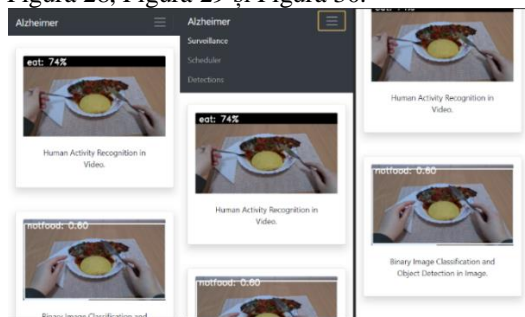


Figura 28. Responsive Surveillance

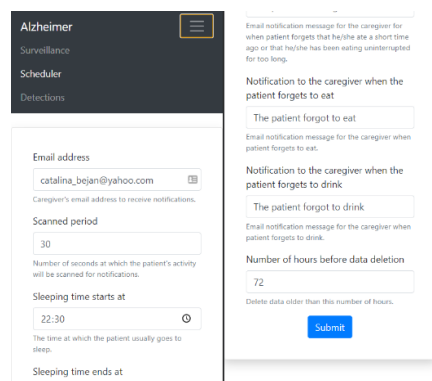


Figura 29. Responsive Scheduler

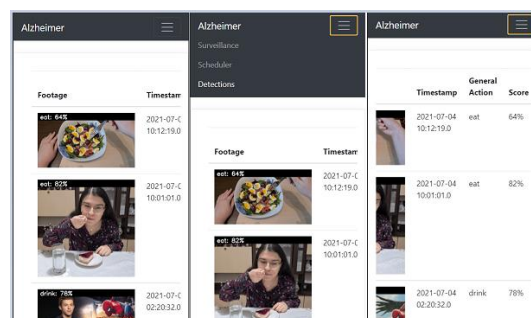


Figura 30. Responsive Detections

Notificările care semnalează momentele de interes înregistrate de sistemul de supraveghere sunt primite de către îngrijitor prin email ca în Figura 31.

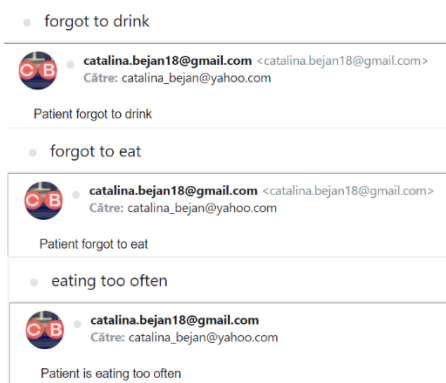


Figura 31. Notificări email

#### 4. CONCLUZIE

Aplicația bazată pe vedere artificială implementată în cadrul acestei lucrări are scopul de a ajuta persoanele bolnave de Alzheimer să își recapete independența, dar și pe îngrijitorii acestora prin ușurarea atribuțiilor lor. Astfel, sistemul permite supravegherea video a bolnavului de Alzheimer și prin intermediul inteligenței artificiale se identifică momentele în care persoana supravegheată se alimentează sau se hidratează.

Sistemul este conceput pentru a preveni subalimentația, supraalimentația sau deshidratarea unei persoane bolnave de Alzheimer prin notificarea acestora printr-un mesaj sonor și reamintirea acțiunilor pe care trebuie să le facă pentru un stil de viață sănătos.

Necesitatea dezvoltării unui astfel de sistem este strict legată de faptul că numărul persoanelor bolnave de Alzheimer este în continuă creștere și deocamdată nu există niciun sistem de supraveghere care să le recunoască acțiunile și să îi ajute să trăiască în mod independent.

Aplicația se remarcă prin conceptul nou pe care îl introduce, recunoașterea activității umane în scopul ajutorării persoanelor bolnave de Alzheimer sau demență și pe îngrijitorii lor.

Sistemul este conceput pentru recunoașterea activității din mai multe perspective: cea în care dispozitivul de supraveghere se află într-o poziție fixă și cea în care dispozitivul este mobil și este purtat de către bolnav.

Modalitatea prin care un astfel de sistem poate fi îmbunătățit este crearea unui set de date suficient de mare și specific acțiunilor de interes pentru antrenarea modelelor.

În viitor se dorește introducerea mai multor roluri de utilizatori care pot accesa aplicația Web și posibilitatea logării acestora prin crearea de conturi în urma cumpărării dispozitivului de supraveghere. Totodată se prevede crearea unei aplicații mobile.

Aplicația va permite conectarea bolnavului de Alzheimer prin identificare biometrică precum amprentă digitală, chip, voce, iris (astfel se evită riscul iminent ca acesta să își uite credențialele de autentificare). Conectarea bolnavului la aplicație va avea ca scop principal amintirea momentului în care acesta a mâncat sau a băut ultima dată printr-o înregistrare video pe care deocamdată o poate accesa numai îngrijitorul.

De asemenea se dorește introducerea rolului de medic în aplicație care să poată gestiona din punct de vedere medical starea de sănătate a pacientului. Introducerea unor statistici ale activităților înregistrate va fi de mare ajutor medicului pentru eventuala stabilire a unui regim alimentar în funcție de obiceiurile pacientului.

Va exista posibilitatea introducerii unui regim alimentar prestabilit de un medic prin identificarea alimentelor pe care le mănâncă (se vor utiliza inclusiv rezultatele detecțiilor de obiecte și clasificare a imaginilor).

În concluzie, sistemul este unul inedit, dar care poate fi de mare ajutor persoanelor bolnave de Alzheimer sau demență, dar și îngrijitorilor acestora prin intermediul inteligenței artificiale.

## 5. BIBLIOGRAFIE

- [1] **S. Gollapudi**, *Learn Computer Vision Using OpenCV: With Deep Learning CNNs and RNNs*, Apress; 1st ed. edition (April 27, 2019), 2019.
- [2] **R. Hartley**, *Multiple View Geometry in Computer Vision* Second Edition, Cambridge University Press, 2004
- [3] **K.-F. Lee**, *AI Superpowers: China, Silicon Valley, and the New World Order*, Houghton Mifflin Harcourt Company; 1st edition (September 1, 2018), 2018
- [4] **S. J. Prince**, *Computer Vision: Models, Learning, and Inference*, Cambridge University Press, 2012.
- [5] **M. Scarpino**, *TensorFlow For Dummies*, For Dummies, April 2018.
- [6] "Activity recognition," Wikipedia, 30 04 2021. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Activity\\_recognition&oldid=1020693189](https://en.wikipedia.org/w/index.php?title=Activity_recognition&oldid=1020693189). [Accessed 16 05 2021].
- [7] "AlexNet," Wikipedia, 06 04 2021. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=AlexNet&oldid=1016307562>. [Accessed 06 04 2021].
- [8] "Alzheimer | Reginamaria.ro," [Online]. Available: <https://www.reginamaria.ro/utile/dictionar-de-afectiuni/alzheimer>. [Accessed 14 03 2021].
- [9] "Analiza-situatie-Alzheimer-2019.pdf," [Online]. Available: <https://insp.gov.ro/sites/cnepss/wp-content/uploads/2017/04/Analiza-situatie-Alzheimer-2019.pdf>. [Accessed 14 03 2021].
- [10] "Apps for people with Alzheimer's or dementia," Alzheimer, [Online]. Available: <https://alzheimer.ie/get-support/equipment-adaptions-and-technology/apps-for-people-with-alzheimers-or-dementia/>. [Accessed 15 03 2021].
- [11] "Baterie externa ASUS ZenPower 10050mAh, 1xUSB, Black," [Online]. Available: [https://altex.ro/baterie-externa-asus-zenpower-10050mah-1xusb-black/cpd/AISBBT026/?gclid=Cj0KCQjw8vqGBhC\\_ARIsADMSd1CB1\\_kA2qPpx7EJE0ES1LetEA1dUshAyjeeY-Nhdv6qoacTsVj9-DQaAvPgEALw\\_wcB](https://altex.ro/baterie-externa-asus-zenpower-10050mah-1xusb-black/cpd/AISBBT026/?gclid=Cj0KCQjw8vqGBhC_ARIsADMSd1CB1_kA2qPpx7EJE0ES1LetEA1dUshAyjeeY-Nhdv6qoacTsVj9-DQaAvPgEALw_wcB). [Accessed 02 07 2021].
- [12] "Boxa portabila PHILIPS, BT25G/00, 2W, Bluetooth, Rosu - eMAG.ro," [Online]. Available: <https://www.emag.ro/boxa-portabila-philips-2w-bluetooth-rosu-bt25g-00/pd/D2ZQV7BBM/>. [Accessed 02 07 2021].
- [13] **J. Carreira and A. Zisserman**, "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset," 12 02 2018. [Online]. Available: <http://arxiv.org/abs/1705.07750>. [Accessed 15 05 2021].
- [14] "Classification: True vs. False and Positive vs. Negative," Google Developers, [Online]. Available: <https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative?hl=ro>. [Accessed 30 06 2021].
- [15] darrenl, "tzutalin/labelImg," GitHub, [Online]. Available: <https://github.com/tzutalin/labelImg>. [Accessed 30 06 2021].
- [16] "Dementia: Technology to stay independent - SCIE," [Online]. Available: <https://www.scie.org.uk/dementia/support/technology/stay-independent>. [Accessed 15 03 2021].
- [17] "Există o legătură între deshidratare și diagnosticul de demență?," [Online]. Available: <https://doc.ro/sanatare/exista-o-legatura-intre-deshidratare-si-diagnosticul-de-dementia>. [Accessed 01 05 2021].
- [18] "Fig. 1 The architecture of ResNet50 and deep learning model flowchart....," ResearchGate, [Online]. Available: [https://www.researchgate.net/figure/The-architecture-of-ResNet50-and-deep-learning-model-flowchart-a-b-Architecture-of\\_fig1\\_334767096](https://www.researchgate.net/figure/The-architecture-of-ResNet50-and-deep-learning-model-flowchart-a-b-Architecture-of_fig1_334767096). [Accessed 30 06 2021].
- [19] "First-person View Of A Guy Eating Food In A Summer Cafe.," Freepik, [Online]. Available: [https://www.freepik.com/premium-photo/first-person-view-guy-eating-food-summer-cafe\\_7862921.htm](https://www.freepik.com/premium-photo/first-person-view-guy-eating-food-summer-cafe_7862921.htm). [Accessed 30 06 2021].
- [20] "Food Images (Food-101)," Kaggle, [Online]. Available: <https://kaggle.com/kmader/food41>. [Accessed 09 04 2021].
- [21] "Fruit Images for Object Detection," [Online]. Available: <https://kaggle.com/mbkinaci/fruit-images-for-object-detection>. [Accessed 30 06 2021].
- [22] "Furniture Detector," Kaggle, [Online]. Available: <https://kaggle.com/akithetechie/furniture-detector>. [Accessed 20 02 2021].
- [23] **K. Hara, H. Kataoka and Y. Satoh**, "Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?," 01 04 2018. [Online]. Available: <http://arxiv.org/abs/1711.09577>. [Accessed 15 05 2021].
- [24] "IMDB-WIKI - 500k+ face images with age and gender labels," [Online]. Available: <https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/>. [Accessed 21 02 2021].
- [25] "In cautarea memoriei. Lupta cu boala Alzheimer," [Online]. Available: <https://www.librarie.net/p/316909/in-cautarea-memoriei-lupta-cu-boala-alzheimer>. [Accessed 14 03 2021].
- [26] **W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, M. Suleyman and A. Zisserman**, "The Kinetics Human Action Video Dataset," 19 05 2017. [Online]. Available: <http://arxiv.org/abs/1705.06950>. [Accessed 15 05 2021].
- [27] **R. Kwok**, "rileykwok/Food-Classification," 30 03 2021. [Online]. Available: <https://github.com/rileykwok/Food-Classification>. [Accessed 09 04 2021].
- [28] **V. Maskara**, "Literature Survey: Human Action Recognition," Medium, 13 03 2021. [Online]. Available: <https://towardsdatascience.com/literature-survey-human-action-recognition-cc7c3818a99a>. [Accessed 15 05 2021].
- [29] "Modified Resnet-34 Architecture," ResearchGate, [Online]. Available: [https://www.researchgate.net/figure/Modified-Resnet-34-Architecture\\_tbl2\\_332522436](https://www.researchgate.net/figure/Modified-Resnet-34-Architecture_tbl2_332522436). [Accessed 01 07 2021].
- [30] **E. Morey**, "What You Need to Know About Dementia and Eating Disorders," The Alzheimer's Site News, 18 01 2017. [Online]. Available: <https://blog.thealzheimerssite.greatergood.com/eating-disorders/>. [Accessed 14 03 2021].

- [31] "opensource.google," [Online]. Available: <https://opensource.google/projects/>. [Accessed 30 06 2021].
- [32] **S. Ren, K. He, R. Girshick and J. Sun**, "*Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*," arXiv.org, 06 01 2016. [Online]. Available: <http://arxiv.org/abs/1506.01497>. [Accessed 30 06 2021].
- [33] **R. Rothe, R. Timofte and L. Van Gool**, "*Deep expectation of real and apparent age from a single image without facial landmarks*," International Journal of Computer Vision, vol. 126, pp. 144-157, 2018.
- [34] **R. Rothe, R. Timofte and L. Van Gool**, "*DEX: Deep EXpectation of apparent age from a single image*," 2015.
- [35] T. R. P. Foundation, "*Buy a Raspberry Pi*," Raspberry Pi, [Online]. Available: <https://www.raspberrypi.org/products/>. [Accessed 02 07 2021].
- [36] "*TensorFlow Hub*," [Online]. Available: <https://tfhub.dev/deepmind/i3d-kinetics-400/1>. [Accessed 15 05 2021].
- [37] "*VGG16 - Convolutional Network for Classification and Detection*," 20 11 2018. [Online]. Available: <https://neurohive.io/en/popular-networks/vgg16/>. [Accessed 28 06 2021]