Universitatea Națională de Știință și Tehnologie POLITEHNICA București

Școala Doctorală de Electronică, Telecomunicații și Tehnologia Informației

Decizie nr. 203 din 21-09-2024

REZUMAT TEZĂ DE DOCTORAT

Ing. Radu-Daniel BOLCAȘ

Contribuţii la recunoaşterea emoţiilor utilizând inteligenţa artificială

COMISIA DE DOCTORAT

Prof.dr.ing. Ion MARGHESCU UNSTPB	Preşedinte
Prof. Dr. Ing. Mihai CIUC UNSTPB	Conducător de doctorat
Prof.dr.ing. Dan-Marius DOBREA Universitatea Tehnică „Gh. Asachi” din Iaşi.	Referent
Prof. Dr. ing. Laurenţiu-Mihail IVANOVICI Universitatea Transilvania din Braşov	Referent
Conf.dr.ing. Eduard POPOVICI UNSTPB	Referent

BUCUREŞTI 2024

Cuprins

Capitolul 1

Introducere

1.1 Prezentarea domeniului tezei de doctorat

Teza de doctorat abordează un studiu din domeniul psihologic și al inteligenței artificiale: recunoașterea automată a emoțiilor atât din expresii faciale observate în imagini, cât și din text.

Recunoașterea emoțiilor folosind Inteligența Artificială (IA) a devenit un domeniu de interes major în ultimii ani, datorită avansurilor în tehnicile de procesare a imaginii, machine learning și deep learning. Aceste tehnologii permit identificarea și clasificarea emoțiilor umane prin analiza expresiilor faciale, a vocii, a gesturilor și chiar a textului scris. În esență, recunoașterea emoțiilor se referă la procesul de detectare și interpretare a semnalelor emoționale exprimate de o persoană, fie că sunt vizuale, auditive sau lingvistice, și are aplicabilitate în domenii diverse, cum ar fi sănătatea, educația, serviciile pentru clienți și securitatea.

Un aspect important al recunoașterii emoțiilor este analiza expresiilor faciale, unde tehnologiile CNN sunt adesea utilizate pentru a identifica trăsături faciale specifice care reflectă stări emoționale. Similar, recunoașterea emoțiilor din voce implică analiza spectrogramelor și utilizarea rețelelor RNN sau Long short-term memory (LSTM) pentru a captura dinamica temporală a vorbirii, fiind aplicabilă în contexte precum centrele de apel sau asistenții virtuali. De asemenea, recunoașterea emoțiilor din text este facilitată de NLP combinat cu modele de deep learning, utilizate pentru analiza sentimentelor pe rețelele sociale sau pentru evaluarea feedback-ului clienților.

1.2 Scopul tezei de doctorat

Aceasta lucrare prezintă cercetări privind recunoașterea emoțiilor faciale și modul în care acestea pot fi folosite în practică. Nevoia de comunicare interumană, de a cunoaște și de a înțelege, duce acest studiu către multiple aplicații în domeniul medical, al securității, al psihologiei, educației, etc.

În cazul modelelor multimodal, acestea pot fi aplicate în diverse domenii, inclusiv diagnosticul medical, psihologia etc. De exemplu, analizarea răspunsurilor scrise ale unei persoane împreună cu expresiile faciale poate oferi o perspectivă valoroasă în starea lor mentală.

Recunoașterea emoțiilor în educație poate aduce numeroase beneficii, contribuind la îmbunătățirea experienței de învățare și la susținerea dezvoltării

emoționale a elevilor. De asemenea se poate identifica stresul sau anxietatea din timp pentru a preveni o posibilă agravare a problemelor emoționale.

1.3 Conţinutul tezei de doctorat

În această teză de doctorat se explorează detectarea automată a emoțiilor din imagini (care conțin expresii faciale) și din text. Teza este organizată în șapte capitole distincte și propune o abordare interdisciplinară care îmbină domeniul psihologiei cu tehnologiile de inteligență artificială.

Capitolul 2 oferă o descriere detaliată a domeniului psihologiei și a modurilor în care se poate realiza recunoașterea emoțiilor, atât prin analiza expresiilor faciale din imagini, cât și clasificarea emoțiilor din text. De asemenea, acest capitol introduce domeniul inteligenței artificiale, prezentând algoritmii și metodele esențiale pentru învățarea automată și profundă, care vor fi utilizate în capitolele următoare.

Capitolul 3 prezintă stadiul actual al recunoașterii emoțiilor faciale în contextul învățării automate, analizând metodele și tehnicile folosite în literatura de specialitate, evidențiind atât avantajele, cât și provocările asociate acestora.

Capitolul 4 explorează utilizarea filtrelor de imagine pentru a accelera sarcinile de procesare a imaginilor. Prin aplicarea filtrelor, scopul este de a accelera procesarea de imagini păstrând în același timp acuratețea modelului. Capitolul cercetează diverse tehnici de filtrare, luând în considerare efectele acestora asupra vitezei de procesare.

Capitolul 5 investighează utilizarea ChatGPT (Generative Pre-trained Transformer) în domeniul recunoașterii emoțiilor faciale, eficientizând procesele de dezvoltare, făcându-le mai ușoare și mai rapide. Cu ChatGPT, scrierea de cod și depanarea se realizează rapid, ceea ce duce la crearea rapidă a unor modele performante în doar câteva minute. Aceast capitol arată, de asemenea, importanța alegerii cuvintelor și a abilităților dezvoltatorului în găsirea echilibrului corect între viteză și precizie.

Capitolul 6 explorează dezvoltarea unui model avansat de recunoaștere a sentimentelor utilizând învățarea multimodală, integrând date text și imagini. Modelul și setul de date multimodal propuse urmăresc să ofere o perspectivă nouă asupra clasificării simultane a datelor text și a imaginilor. Studiul evidențiază beneficiile învățării multimodale în gestionarea informațiilor ambigue și îmbunătățirea sarcinilor de clasificare.

În ultimul capitol 7, se formulează concluziile tezei, rezumând rezultatele obținute și contribuțiile proprii aduse. Totodată, sunt analizate implicațiile acestor rezultate și se sugerează posibile direcții pentru cercetările viitoare.

Capitolul 2

Prezentare generală

2.1 Recunoașterea emoțiilor

Comunicarea interpersonală a fost întotdeauna esențială pentru oameni, iar abilitățile umane de a interpreta starea de spirit a altora pot duce la erori. Emoțiile sunt procese mentale declanșate de stimuli interni sau externi. Deși oamenii pot învăța să-și disimuleze gesturile, expresiile faciale pot dezvălui emoții autentice, iar tehnologia poate ajuta la detectarea acestora [1]. În literatura de specialitate se definesc un număr limitat de emoții primare, universale indiferent de rasă, societate sau cultură [2].

2.2 Inteligența Artificială, Învățarea Automată (machine learning) și Învățarea Profundă (Deep learning)

2.2.1 Inteligența artificială

Inteligența artificială (IA) este un domeniu al informaticii dedicat dezvoltării de mașini și software capabile să îndeplinească sarcini ce necesită inteligență umană. Există două categorii principale de IA: IA slabă și IA puternică.

2.2.2 Învățarea Automată (Machine Learning)

Învățarea automată (ML) este un subdomeniu al IA în care computerele pot învăța din date și să generalizeze pentru cazuri necunoscute. Există trei tipuri principale de învățare automată: învățarea supravegheată, învățarea nesupravegheată, învățarea prin consolidare.

2.2.3 Învățarea profundă (Deep Learning)

Învățarea profundă (DL) este un subdomeniu al ML care folosește rețele neuronale artificiale complexe pentru a învăța și a face predicții din date. Acest proces permite modelului să învețe reprezentări Abstracte. Tehnicile utilizate includ rețele

neuronale convoluționale (CNN), rețele neuronale recurente (RNN) și rețele Transformer (BERT [3], GPT [4]).

2.3 Rețele Neurale Artificiale

În DL, modelul utilizează rețele neuronale artificiale cu multiple straturi. Aceste rețele își ajustează parametrii printr-un mecanism de feedback, utilizând funcția de cost pentru a evalua corectitudinea rezultatelor obținute. Funcțiile de activare, cum ar fi ReLU, transformă datele dintr-un strat pentru a le transmite la următorul, în timp ce optimizatorii, ajută la minimizarea funcțiilor de eroare și la ajustarea ponderilor modelului.

2.4 Rețele Neuronale Convoluționale (CNN)

CNN sunt folosite pentru a procesa imagini, iar în acest context, imagini ce conțin expresii faciale. Straturile de convoluție sunt utilizate pentru a izola caracteristici precum sprâncenele, nasul și gura de alte elemente nerelevante. Acest proces permite sistemului să clasifice expresiile faciale în mod eficient.

2.4.1 Straturile Convoluționale

Straturile convoluționale procesează imagini, recunosc și extrag caracteristicile relevante. Aceste straturi utilizează kerneluri aplică un produs scalar asupra imaginii, iar matricea rezultată este trimisă către următorul strat de neuroni. Neuronii primesc informații de la stratul anterior, aplică o funcție matematică și transmit rezultatul mai departe.

2.4.2 Stratul de extracție (Pooling)

Stratul de extracție este în general utilizat după un strat de convoluție și are ca scop reducerea datelor, dar și extragerea celor mai importante caracteristici printr-o fereastră ce se va deplasa, calculând cu fiecare deplasare maximul/medierea valorilor pe care fereastra le conține.

2.5 Descrierea hardware și software

Implementarea a fost realizată pe o arhitectură bazată pe Linux și Docker, folosind Ubuntu 22.04 LTS și platformele TensorFlow și Jupyter. Antrenarea modelelor s-a efectuat utilizând o placă grafică NVIDIA GeForce GTX 1070.

Capitolul 3

Recunoașterea emoțiilor faciale in învățarea automată [1]

Acest capitol este tradus și adaptat după articolul autorului cu numele „Facial Emotions Recognition in Machine Learning” [1].

Aceast capitol conține o descriere a aspectelor psihologice ale FER și oferă o descriere a seturilor de date și algoritmilor care fac posibile rețelele neuronale. Se realizează și o revizuire a literaturii asupra studiilor recente în recunoașterea emoțiilor faciale, detaliind metodele și algoritmii utilizați pentru a îmbunătăți capacitățile sistemelor care folosesc învățarea automată. Sunt discutate provocările legate de învățarea automată, cum ar fi supraînvățarea, posibilele cauze, soluții și provocările legate de seturile de date, cum ar fi discrepanțele ca, orientarea capului, iluminarea, și dezechilibrul claselor în setul de date.

3.1 Introducere

Comunitatea psihologică definește un număr mic de emoții de bază care sunt exprimate la fel, indiferent de rasă, gen, origine, societate sau regiune geografică de unde provin [2]. Emoțiile de bază sunt furie, fericire, dezgust, surpriză, tristețe, dispreț și frică.

Pe lângă aceste emoții de bază, un alt model utilizat pentru a determina emoțiile este Sistemul de Codificare a Acțiunilor Faciale (FACS). Acesta a fost dezvoltat de Ekman și Friesen și constă din coeficienți numiți Unități de Acțiune (AU), care pot defini majoritatea expresiilor faciale posibile [5]. Prin combinarea diferitelor valori AU, este posibilă determinarea emoției la un moment specific.

3.2 Algoritmi și seturi de date

Metodologii diferite au fost utilizate de cercetători, de la învățarea automată clasică la învățarea profundă. Învățarea automată clasică se bazează pe primii pași de

a prelucra datele într-un mod care să extragă caracteristicile relevante ale imaginii și apoi să ofere aceste caracteristici ca intrare pentru un clasificator.

Învățarea profundă este un subset al învățării automate și este concepută pentru a minimiza intervenția umană. Din această cauză, tinde să fie mai complexă și necesită mai multe resurse hardware decât abordarea clasică, oferind în același timp rezultate mai bune.

Rețelele neuronale convoluționale (CNN) sunt o alegere populară de rețele neuronale artificiale profunde capabile să determine tipare din datele de intrare. Cea mai utilizată aplicație este pentru problemele de analiză și clasificare a datelor de imagine, dar clasificarea videoclipurilor sau a textelor este, de asemenea, posibilă. Acestea pot lucra cu tipuri de date 1D sau 3D, în funcție de imagini, dacă sunt în nuanțe de gri sau colorate. Este proiectată după modelul creierului uman și încearcă să imite același proces de învățare.

Bazele de date sunt importante atunci când discutăm despre FER. Ele reprezintă un factor crucial care determină acuratețea recunoașterii. Bazele de date conțin un număr diferit de imagini sau videoclipuri și au caracteristici specifice [6]. Unele dintre aceste caracteristici includ înclinarea și orientarea feței, iluminarea, numărul de actori care performează, contextul în care emoțiile au fost înregistrate (de exemplu, condiții de laborator etc.), accesorii sau părul facial și, desigur, numărul de emoții.

Baza de date FER2013 [7] este o colecție la scară largă de imagini obținute cu Google image search API. Imaginile au fost procesate la 48*48 pixeli și redimensionate. Conține 35.887 de imagini în tonuri de gri cu șase emoții de bază și neutru.

3.3 Analiza literaturii de specialitate

Cercetătorii lucrează pentru a îmbunătăți capabilitățile sistemelor FER fie prin creșterea preciziei, fie prin reducerea timpului de antrenare sau a puterii de procesare necesare pentru antrenare, găsind modalități de a utiliza datele mai eficient. Deși acestea nu sunt singurele căi, ele sunt cele mai cercetate.

Într-o abordare privind detectarea emoțiilor faciale macro, Rzayeva și Alasgarov [8] au utilizat bazele de date CK+ și RAVDESS. După colectarea și preprocesarea datelor, a fost utilizată o rețea CNN. În 50 de epoci, precizia a atins 80%. A doua lor abordare a folosit VGG16 [9] și a dus la o precizie de 82%, dar și la suprainvățare semnificativă. Pentru a rezolva această problemă, au fost introduse straturi dropout. Modelul propus a avut precizia la 88% pentru CK+, 92% pentru RAVDESS și 92% pentru bazele de date CK+ și RAVDESS.

Într-un studiu realizat de Tarnowski și colab. [10], AU au fost utilizate pentru a determina caracteristicile expresiilor faciale folosind un model facial tridimensional. Metodologiile utilizate sunt MLP cu clasificatorul k-NN. Utilizarea Microsoft Kinect pentru a modela o față 3D și pentru a cartografia punctele de pe o față pe marginile

trăsăturilor faciale. Punctele au fost utilizate împreună cu FACS [5] pentru a determina caracteristicile faciale (colțul gurii, sprâncenele, nasul, pomeții). Precizia finală a fost de 96% pentru 3-KNN și 90% pentru MLP.

Într-un studiu realizat de Yang și colab. [11], a fost propusă o nouă abordare pentru FER folosind o componentă expresivă și o componentă neutră. Cercetând literatura psihologică, o expresie facială a fost descompusă într-o componentă expresivă și una neutră. Rețeaua primește la intrare 2 poze cu subiectul într-o stare neutră și o poză cu o anumită emoție. Modelul creează o a doua imagine a aceleași persoane într-o stare neutră. Rețeaua CNN propusă a fost pre-antrenată pe Binghamton University 3D Facial Expression (BU-3DFE) și BP4D-Spontaneous și a reușit să atingă o acuratețe de 75.23% pe baza de date MMI, 88% pe setul Oulu-CASIA și 97.30% pe baza de date CK+.

3.4 Analiza cercetărilor

Printre provocările observate se numără dezechilibrul claselor, care poate crea un bias în predicție, actori instruiți să exprime o anumită emoție ce poate devia de la o expresie naturală, precum și expresiile cu un nivel scăzut de intensitate. De asemenea, un număr mic de subiecți din setul de date poate conduce la supraînvățare dacă imaginile sunt prea asemănătoare. Dificultăți suplimentare apar în recunoașterea fețelor înclinate sau rotite, precum și a imaginilor cu iluminare slabă, umbre și contrast. Subiecții care poartă ochelari, au păr facial sau haine care obstrucționează trăsături importante ale feței pot reprezenta alte obstacole. Este, de asemenea, importantă evitarea supraînvățării și subînvățării, deoarece acestea afectează eficiența rețelei în determinarea unui optim. În plus, timpul necesar pentru a antrena o rețea CNN este, în general, mare din cauza numărului considerabil de straturi și neuroni, ceea ce face procesarea intensă.

3.5 Concluzii

Studiile arată că modelele cresc în acurateţe fără a deveni neapărat mai complexe. Pentru modelele simple au fost utilizate abordările care măsoară distanța față de caracteristicile feței sau maparea vederii 3D a feței combinate cu FACS. Această mapare a feței a reușit să depășească simplitatea modelului, având în același timp o acuratețe ridicată.

O abordare interesantă în FER a fost realizată și prin utilizarea unei componente expresive și a unei componente neutre [11]. Această metodă a dat rezultate bune, deși modelul este complex și poate avea nevoie de un set de date și mai mare pentru a continua extinderea.

În acest capitol au fost evaluate provocările în dezvoltarea unui sistem FER pentru a oferi o perspectivă asupra direcțiilor potențiale de cercetare ce pot fi luate în considerare.

În cadrul articolului realizat de Diana Dranga și **Radu-Daniel Bolcaş** [12], s-a analizat o posibilă implementare a unui model text de rețea neuronală convoluțională pentru a ușura depanarea circuitelor digitale în domeniul verificării funcționale. Acest articol conține o revizuire a modului în care Inteligența Artificială poate reduce acest blocaj, luând în considerare timpul petrecut pentru implementarea mediului de verificare și timpul necesar pentru atingerea procentajului dorit de acoperire.

Au fost evidențiate realizările, provocările și tehnicile de învățare automată precum CNN, RNN, etc. pentru domeniul software sau hardware.

O posibilă direcție de cercetare reprezintă analiza obiectivității cerințelor ("requirements"), se pot adopta mai multe strategii. În mod uzual documentele tehnice sunt scrie într-un limbaj neutru din punct de vedere al emoțiilor.

În domeniul testării cu cât o problemă este descoperită mai târziu, cu atât este mai costisitoare, deoarece după ce se realizează schimbările necesare, se reiau toți pașii implicând mulți ingineri ce provoacă consturi și o pierdere de timp. Această analiză poate crea o nouă prioritizarea a cerințelor în testare pentru a avea o validare inițială urmată de o validare cuprinzătoare. În această abordare, eventualele probleme majore pot fi detectate cât mai repede în procesul de dezvoltare și verificare. Această direcție de cercetare este promițătoare și va fi cercetată în lucrări și articole viitoare.

Capitolul 4

Îmbunătățirea eficienței antrenării în recunoașterea emoțiilor faciale [13]

Acest capitol este tradus și adaptat dupa articolul autorului cu numele „Enhancing Training Efficiency in Facial Emotion Recognition” [13].

În cadrul acestui capitol se explorează utilizarea filtrelor de imagine pentru a accelera sarcinile de procesare a imaginilor. Prin aplicarea filtrelor, scopul este de a accelera procesarea de imagini păstrând în același timp acuratețea modelului.

4.1 Introduction

Rețelele neuronale convoluționale (CNN) au demonstrat că au potențialul de a aduce rezultate bune în FER. Deși performanța este excelentă, procesul de antrenare este consumator de timp.

O soluție pentru a îmbunătăți atât acuratețea, cât și timpul de antrenare al modelului este filtrarea informațiilor din setul de date înainte de a le introduce. O idee este să se determine trăsăturile faciale dintr-o imagine și apoi să se utilizeze această imagine pentru a începe antrenarea. În acest fel, toate caracteristicile nerelevante sunt filtrate.

Aceast capitol prezintă o abordare pentru reducerea cantității de caracteristici nerelevante, accelerând în același timp procesul de antrenare. Pentru a crește performanța și pentru a încerca îmbunătățirea acurateții pentru un model specific (dezvoltat de Y. Khaireddin și Z. Chen [14]), a fost adăugat un pas suplimentar de preprocesare care constă în filtre pentru a îmbunătăți datele trimise modelului. Mai mult, folosind această nouă abordare, timpul necesar pentru antrenarea unui model a scăzut semnificativ.

4.2 Implementarea

4.2.1 Prezentare generală

Utilizarea filtrelor de imagine a devenit o tehnică puternică, transformând modul în care CNN percep și manipulează conținutul vizual. În domeniul viziunii computerizate, aplicarea filtrelor de imagine a apărut ca un instrument puternic, capabil să elimine fără probleme informațiile nedorite din imagini. Prezența elementelor nedorite, cum ar fi zgomotul, obiectele și fundalul, poate compromite semnificativ calitatea caracteristicilor pe care o CNN le poate extrage dintr-o imagine.

4.2.2 Metode de preprocesare

Tehnicile de preprocesare aplicate implică utilizarea a patru filtre: un filtru de prag, două filtre adaptive de prag și metoda lui Otsu. Aceste filtre sunt incluse în biblioteca OpenCV [15]. OpenCV este un software recunoscut pe scară largă, utilizat pentru procesarea imaginilor și a videoclipurilor, care cuprinde o multitudine de funcționalități.

Primul filtru este un simplu filtru de prag. O valoare uniformă a pragului este utilizată pentru fiecare pixel. Pixelii cu valori sub prag sunt setați la 0, în timp ce cei care îl depășesc sunt setați la o valoare maximă predefinită [16].

Al doilea și al treilea filtru aparțin categoriei filtrelor adaptive de prag. Algoritmul utilizat de acest filtru calculează pragul luând în considerare regiunea locală din jurul fiecărui pixel. Al doilea filtru calculează valoarea pragului prin media aritmetică a zonei din vecinătate, scăzând și constanta C [15,16]. Al treilea filtru folosește „o sumă ponderată gaussiană a valorilor din vecinătate minus constanta C” [15].

Al patrulea filtru implică o determinare automată a pragului utilizând imaginea însăși. Această metodă denumită metoda lui Otsu, utilizează histograma imaginii pentru a identifica vârfurile din grafic și selectează o valoare situată între aceste vârfuri [16]. Avantajul în acest caz este nevoia minimă de ajustare a parametrilor pentru procesul de filtrare.

4.2.3 Arhitectura CNN și preprocesarea datelor

Într-un studiu din 2021 realizat de Y. Khaireddin și Z. Chen, setul de date FER2013 a fost utilizat pentru a obține performanțe ridicate cu o rețea de dimensiuni medii. Ei au folosit un model de rețea neuronală convoluțională și au ajustat fin hiperparametrii acesteia [14].

4.3 Setul de date

FER2013 [7] constă din 35.888 de imagini reprezentând 7 emoții diferite. Aceste imagini sunt împărțite în 3 categorii: antrenare, validare și testare. Această împărțire a fost făcută la momentul publicării în ICML. În această lucrare, împărțirea pentru antrenare, validare și testare a fost realizată diferit, adăugând mai multe imagini în datele de antrenare, obținând astfel o îmbunătățire a timpului de antrenare. Din acest motiv, acuratețea afișată poate prezenta o variație comparativ cu literatura de specialitate.

4.4 Rezultate și analiză

Prin aplicarea adaptării pentru baza de date FER2013 s-a obținut un rezultat de 65,0124%. Timpul necesar pentru o epocă este de 194,0375s. Această valoare va servi ca punct de referință pentru comparație cu filtrele utilizate. Valorile uzuale obținute în literatură de specialitate, variază în jurul valorii de 67% +/- 5%. Această scădere se datorează modului în care sunt structurate datele.

Aplicarea filtrului simplu de prag a dus la o acuratețe de 57,0452%, obținută după rularea a 30 de epoci. Timpul necesar pentru o epocă este de 128,9716s.

Al doilea filtru introdus este filtrul de prag adaptiv de medie. Acest filtru duce la o acuratețe de 56,6406% după 30 de epoci. Timpul necesar pentru o epocă a fost de 129,1821s.

Al treilea filtru utilizat, este filtrul de prag adaptiv gaussian. Prin testele efectuate, cea mai mare acuratețe obținută pentru datele de validare a fost de 62,9883%, apropiindu-se foarte mult de valoarea de referință. Acest rezultat este ilustrat vizual în Figura 4.. Timpul scurs pentru o epocă este de 128,7093s.

Figura 4.7 *Utilizarea filtrului de prag adaptiv gaussian (Adaptive Thresh Gaussian)*

Transformarea vizuală a unei imagini folosind pragul Gaussian este observabilă în Figura 4..

Figura 4.8 *Aplicarea filtrului prag adaptiv gaussian (Adaptive Thresh Gaussian) pe fotografie*

Ultimul filtru utilizat în experimente a fost cel bazat pe metoda lui Otsu. Prin determinarea automată a pragului, acest filtru este capabil să producă o acuratețe de 53,6830% cu un timp de antrenare pe epocă de 128,9333s.

Cel mai important efect al filtrelor este văzut prin cuantificarea timpului necesar pentru antrenarea standard a modelului și compararea acestuia cu utilizarea filtrelor. Comparând abordarea fără filtru cu filtrul adaptiv Gaussian se observă o reducere notabilă de 33,6678% în durata antrenării fiecărei epoci. Prin urmare, antrenarea folosind filtrele poate îmbunătăți semnificativ timpul necesar pentru antrenarea unui model. În special pentru modelele complexe, această abordare poate servi ca un instrument puternic pentru accelerarea procesului de antrenare.

4.5 Concluzii

Prin selectarea și aplicarea atentă a filtrelor, timpul de procesare al imaginilor poate fi redus semnificativ. Experimentele efectuate au evidențiat eficacitatea diferitelor tehnici de filtrare în îmbunătățirea vitezei de procesare cu ~33% fără a compromite acuratețea (o variație de ~2%).

Rezultatele arată importanța selecției filtrelor și optimizării parametrilor pentru a obține compromisurile dorite între viteza de procesare și acuratețe. Informațiile obținute din acest studiu contribuie la o înțelegere mai profundă a modului în care filtrele de imagine pot fi valorificate pentru a procesa imaginile mai rapide și mai eficient.

Capitolul 5

Generarea de modele FER utilizând ChatGPT [17]

Acest capitol este tradus și adaptat după articolul autorului cu numele „Generating FER models using ChatGPT” [17].

Aceast capitol investighează utilizarea ChatGPT în domeniul recunoașterii emoțiilor faciale, eficientizând procesele de dezvoltare, făcându-le mai ușoare și mai rapide. Cu ChatGPT, scrierea de cod și depanarea se realizează rapid, ceea ce duce la crearea rapidă a unor modele performante în doar câteva minute. De asemenea, este importantă alegerea cuvintelor și abilitatea dezvoltatorului în găsirea echilibrului între viteză și precizie.

5.1 Introducere

În domeniul recunoașterii emoțiilor faciale (FER) se operează în principal pe învățarea supravegheată care este abordarea dominantă, în timp ce explorarea învățării nesupravegheate a fost limitată. Dezvoltarea modelelor implică mai multe etape: selectarea unui set de date adecvat, preprocesarea pentru standardizare și pregătire a datelor, alegerea arhitecturii modelului, antrenarea acestuia, ajustarea hiperparametrilor și, în final, evaluarea modelului.

Abordarea propusă utilizează ChatGPT de la OpenAI [4] pentru a genera codul inițial, permițând cercetătorilor să facă ajustări ulterioare. Această utilizare ca instrument de suport reduce semnificativ timpul necesar pentru dezvoltare și validarea inițială. După finalizarea procesului și obținerea rezultatelor, se evaluează dacă modelul este un punct de plecare bun sau este necesară o altă arhitectură.

5.2 Prezentare generală

5.2.1 Modelele lingvistice mari (LLM)

Modelele lingvistice mari (LLM) sunt sisteme de inteligență artificială antrenate pe volume mari de date text pentru a înțelege și genera limbaj uman. Exemple precum ChatGPT, dezvoltat de OpenAI, pot procesa și genera text care imită

îndeaproape limbajul uman. Aceste modele înțeleg contextul, semantica și sintaxa, iar evaluarea lor se concentrează pe abilitatea de a produce text coerent și relevant în contexte specifice.

5.2.2 Arhitectura FER și setul de date

Baza de date utilizată în acest studiu FER2013 [7], este un set de date folosit pe scară largă în comunitatea de cercetare. În recunoașterea emoțiilor faciale, abordările de învățare nesupervizată implică adesea abordări de a grupa expresiile faciale similare sau pentru a extrage caracteristicile relevante din imaginile faciale ce pot fi utile atunci când datele etichetate sunt rare.

5.3 Implementare și rezultate

ChatGPT a fost folosit pentru a ajuta la stabilirea modelului de bază și pentru investigații ulterioare. Formularea întrebărilor poate influența răspunsurile obținute. Prin urmare, a fost adoptată o abordare de a genera codul de bază utilizând o abordare zero-shot. Astfel, ChatGPT a fost provocat să genereze modele utilizând atât învățarea supravegheată, cât și cea nesupravegheată.

5.3.1 ChatGPT cu învățare nesupervizată

Întrebarea inițială adresată lui ChatGPT a fost să "Creeze un model FER folosind învățarea nesupervizată cu setul de date FER2013". A generat un model care integrează StandardScaler pentru standardizarea datelor, PCA pentru reducerea dimensionalității și identificarea componentelor primare și secundare, și K-Means ca algoritm de grupare. Grupurile rezultate sunt ilustrate în Figura 5.1.

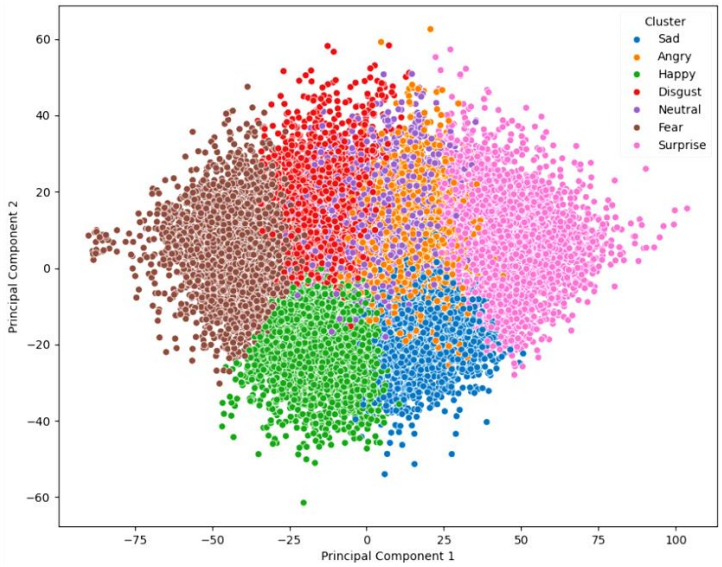

Figura 5.1 *K-Means Cluster pentru FER2013*

O a doua întrebare adresată lui ChatGPT a fost să folosească un alt algoritm care să ofere un model similar, folosind Modele de Amestec Gaussian (GMM) în loc de K-means pentru a grupa etichetele. S-a obținut o performanță sub standard pentru acest algoritm de clustering.

După multiple conversații cu ChatGPT, s-a ajuns la un punct în care erorile și necesitatea de a regenera au devenit și mai pronunțate. Pe măsură ce modelele generate au început să arate rezultate și mai slabe, iar erorile au devenit mai frecvente, s-a ales încheierea aceastei direcții de cercetare și trecerea la abordarea supervizată.

5.3.1 ChatGPT cu învățare supervizată

Întrebarea inițială adresată a fost: "Crează un model de învățare automată pentru recunoașterea emoțiilor în imagini folosind setul de date FER2013". Cu toate că designul său a fost simplist, ChatGPT3.5 a furnizat tot codul necesar pentru a executa modelul conform specificațiilor. Cu toate acestea, metricile obținute au fost suboptime, cu o acuratețe de 51.14%.

Au urmat multiple mesaje unde informațiile primite erau în mare parte teoretice și deși exacte într-un sens general, nu se aplicau direct scenariului specific dorit. Modelele propuse ulterior de ChatGPT au implicat intervenția autorului pentru a le face funcționale. Din cauza numărului mare de conversații cu ChatGPT, detaliile inițiale anterioare au devenit mai puțin pertinente. Prin urmare, mesajele următoare oferă informații cuprinzătoare în fiecare întrebare, ceea ce s-a dovedit mai eficient.

Încercând această abordare, prin multiple regenerări, a apărut un model distinct cu blocuri reziduale. Antrenat timp de 30 de epoci, a atins o acuratețe de 62.49%.

Folosind aceeași abordare, a fost obținut un nou model cu trei blocuri și antrenat timp de 20 de epoci, acuratețea obținută este de 60.08%. Deși acuratețea nu este la fel de mare ca în cazul modelului cu blocuri reziduale, acest model evită problemele de supraînvățare sau subînvățare, așa cum este arătat în Figura 5.6 și Figura 5.7.

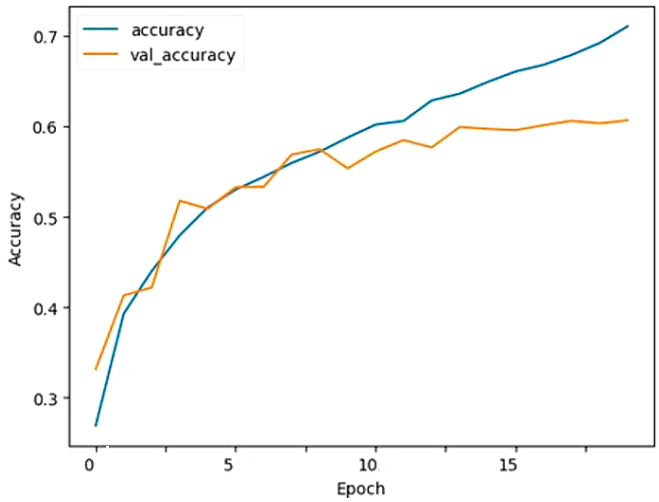

Figura 5.6 *Acuratețea modelului cu trei blocuri antrenat cu 20 de epoci*

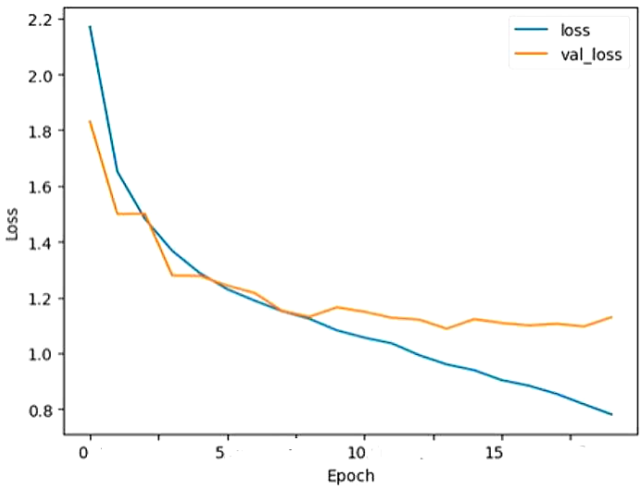

Figura 5.7 *Funcția de pierdere a modelului cu trei blocuri antrenat cu 20 de epoci*

5.4 Analiza rezultatelor

Procesul de lucru cu ChatGPT pentru generarea de modele a evidențiat atât succese, cât și provocări, mai ales în cazul învățării nesupervizate, care a avut o acuratețe scăzută. Diverse strategii, inclusiv modificarea arhitecturilor și preprocesarea datelor, au dus la rezultate variabile și în anumite cazuri, au implicat intervenția autorului pentru a le face funcționale. Totuși, interacțiunile repetate cu ChatGPT au accelerat dezvoltarea, deși calitatea rezultatelor a depins de claritatea mesajelor.

Rezultatele obținute, cu o acuratețe între 60.08% și 62.49% în recunoașterea emoțiilor faciale, evidențiază eficiența rapidă a dezvoltării modelului, deși este ușor sub modelele de ultimă generație (73.28%-75.97%). Principalul avantaj constă în capacitatea de a itera rapid și de a experimenta diverse configurări, făcând această abordare ideală în medii cu termene strânse. Utilizarea ChatGPT a demonstrat potențialul LLM-urilor de a crește eficiența și agilitatea în crearea aplicațiilor de AI pentru recunoașterea emoțiilor faciale.

5.5 Concluzii

Studiul explorează utilizarea ChatGPT pentru accelerarea dezvoltării de modele, reducând semnificativ timpul alocat codării și depanării. Deși în mediul nesupervizat rezultatele au fost mai slabe, în mediul supravegheat s-a obținut o acuratețe de 60.08%, iar arhitectura a fost validată în câteva minute. Cercetătorii au astfel un punct de plecare solid pentru îmbunătățiri ulterioare. Rezultatele subliniază importanța selecției cuvintelor și a expertizei dezvoltatorilor în găsirea echilibrului între viteza de dezvoltare și acuratețe, demonstrând potențialul LLM-urilor de a eficientiza procesele și a deveni un instrument valoros pentru dezvoltatori.

Capitolul 6

Ansamblu de modele pentru analiza multimodală a sentimentelor utilizând fuziunea textelor și a imaginilor [18]

Acest capitol este tradus și adaptat după articolul autorului cu numele „Ensemble models for multimodal sentiment analysis using textual and image fusion” [18].

Analiza sentimentelor este un domeniu în evoluție, care atrage un interes de cercetare semnificativ. Analiza multimodală a sentimentelor (MSA) integrează diferite forme de date, cum ar fi textul pentru recunoașterea emoțiilor și imaginile prin recunoașterea emoțiilor faciale (FER), procesând diverse modalități de intrare. Această lucrare introduce ImaText, un set de date nou pentru recunoașterea emoțiilor care combină texte și imagini din DailyDialog și FER2013. Modelul multimodal propus și setul de date oferă o perspectivă nouă asupra clasificării simultane a datelor text și imagine.

6.1 Introducere

În aceast capitol, principalele două forme de recunoaștere a emoțiilor vor fi recunoașterea emoțiilor faciale (FER) și recunoașterea emoțiilor din text.

Analiza sentimentelor multimodale (MSA) implică integrarea diferitelor forme de date, inclusiv imagini, text, audio sau video, pentru a procesa multiple modalități de intrare sau ieșire. Prin integrarea diverselor modalități, capacitățile modelului pot fi îmbogățite semnificativ.

În contextul învățării multimodale, literatura de specialitate existentă conturează în mod obișnuit două niveluri de fuziune: fuziunea la nivel de caracteristici, adesea denumită fuziune timpurie, și fuziunea la nivel de decizie, cunoscută și sub numele de fuziune târzie.

6.2 Setul de date și preprocesarea lor

Acest nou studiu propune o fuziune între două seturi de date, unul conținând text etichetat cu emoții (DailyDialog [19]) și al doilea conținând imagini etichetate cu emoții faciale (FER2013). Corelând emoțiile din ambele seturi de date, rezultatul constă într-un fișier CSV și o structură de directoare în care sunt localizate imaginile.

Acest nou set de date numit **ImaText** cuprinde 25.780 de intrări pentru șase emoții și a fost salvat pentru învățarea multimodală. Distribuția finală este de 4.865 pentru "furie", 555 pentru "dezgust", 1.255 pentru "frică", 8.910 pentru "fericire", 6.190 pentru "tristețe" și 4.005 pentru "surpriză".

6.3 Arhitectura și modelul propus

Arhitectura aleasă este o rețea neuronală convoluțională (CNN), optimizată pentru analizarea textului și imaginilor. Pentru text, un tokenizer mapează cuvintele la indexuri, urmat de padding pentru uniformizarea lungimii textului. Datele sunt împărțite în imagini, texte și etichete, iar apoi în seturi de antrenament și testare. Modelul textului folosește straturi de intrare, embedding, convoluție unidimensională, funcții de activare (ReLU), dropout pentru prevenirea supraînvățării, max și un strat final complet conectat care clasifică datele în șase clase distincte.

Modelul imaginii începe cu un strat de intrare similar cu cel al modelului de text. Urmează un bloc format dintr-un strat de convoluție bidimensională și un strat de max pooling, care extrag și comprimă caracteristicile imaginii. Acest bloc se repetă de trei ori. Un strat flatten restructurează datele pentru stratul complet conectat, responsabil de clasificarea emoțiilor.

Pentru a dezvolta un model multimodal, ieșirile modelelor de text și imagine sunt combinate printr-un strat de "concatenate". După concatenare, se adaugă două straturi complet conectate, iar stratul final clasifică datele în șase clase de emoții.

6.4 Rezultate și analiză

Modelul multimodal propus a performat bine pe setul de date creat. Au fost efectuate diverse experimente, în timpul cărora modelul a suferit adaptări și îmbunătățiri, inclusiv testarea diferiților optimizatori și ajustarea hiperparametrilor. Modelul multimodal a fost antrenat și a obținut o acuratețe de validare de 70.19%, așa cum este ilustrat în Figura 6.4, în timp ce graficul pierderii este prezentat în Figura 6.5.

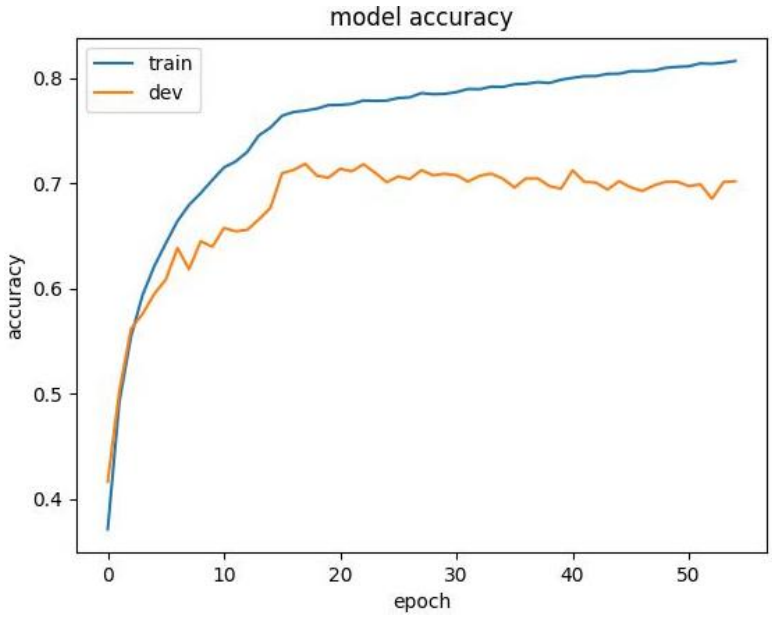

Figura 6.4 *Acuratețea modelului multimodal*

Modelul antrenat cu date atât de imagine, cât și de text, a prezentat ușoare semne de supraantrenare începând cu 15 epoci și a devenit pronunțat începând de la 25 epoci. Performanța optimă a fost observată în jurul epocii 25.

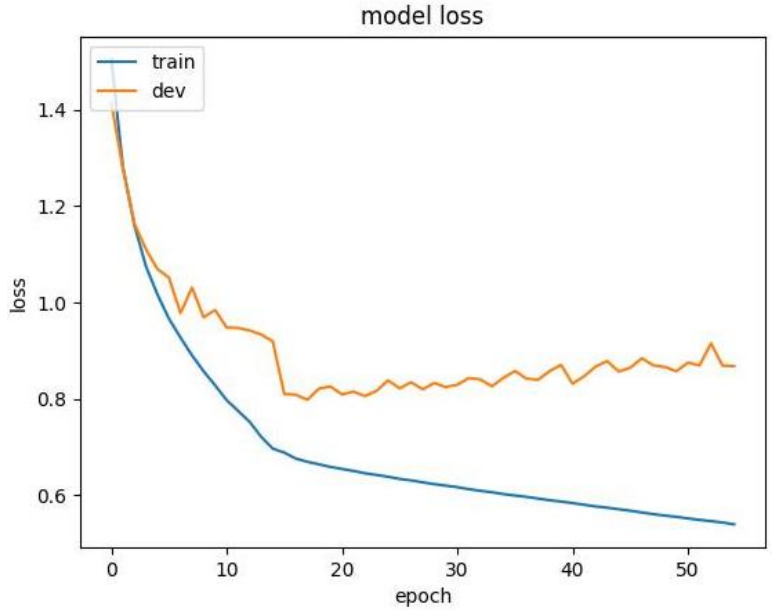

Figura 6.5 *Pierderea modelului multimodal*

Modelul a întâmpinat dificultăți în recunoașterea anumitor emoții din cauza dezechilibrului între clase, deoarece unele emoții reprezentau sub 5% din date, iar modelul a avut tendința de a le ignora. În viitor, se pot folosi tehnici de reducere a acestui dezechilibru, precum augmentarea datelor. Pe setul de date DailyDialog, modelele de clasificare existente au atins aproximativ 59% [20], iar pe FER2013, este în jur de 70% +/- 5% [14,21]. Modelul propus a obținut 70.19% pe noul set de date, evidențiind potențialul abordării multimodale, unde integrarea informațiilor din text și imagini poate îmbunătăți învățarea și performanța.

6.5 Concluzii

Acest capitol prezintă implementarea cu succes a unui nou set de date, ImaText, creat prin combinarea DailyDialog și FER2013, și a unui model multimodal care a obținut o acuratețe de 70,19%. ImaText este primul set de date multimodal creat special pentru analiza sentimentelor, care include doar text și imagini. Acesta conține 25.780 de înregistrări distribuite în șase emoții. Autorul a proiectat și ajustat modelul pentru a se adapta acestui set de date inovator.

Rezultatele sunt inovatoare, deschizând noi direcții de cercetare în analiza multimodală a sentimentelor, cu aplicații în diagnostic medical și îmbunătățirea interacțiunii om-computer. Studiul propune și utilizarea datelor existente pentru a augmenta performanța modelelor, oferind o soluție la provocarea lipsei seturilor de date multimodale.

Capitolul 7

Concluzii

Teza de doctorat investighează recunoașterea automată a emoțiilor folosind algoritmi de învățare automată, concentrându-se pe expresii faciale din imagini și text expresiv. Au fost utilizate baze de date precum FER2013 și DailyDialog, iar autorul a creat o bază de date proprie numită ImaText, care combină imagini și text etichetat cu emoții. Studiul explorează un model avansat de recunoaștere multimodală, care integrează atât textul, cât și imaginile, obținând o acuratețe de 70.19%. De asemenea, lucrarea evaluează tehnici de filtrare pentru a optimiza viteza procesării imaginilor și examinează utilizarea ChatGPT ca un instrument pentru a accelera dezvoltarea modelelor, subliniind importanța selecției cuvintelor și a competențelor dezvoltatorului pentru a obține un echilibru optim între viteză și precizie.

7.1 Rezultate obţinute

În capitolul 3 s-a descris aspectele psihologice ale FER și se oferă o descriere a seturilor de date și algoritmilor care fac posibile rețelele neuronale. Apoi, se realizează o revizuire a literaturii de specialitate asupra studiilor recente în FER. Sunt discutate provocările legate de învățarea automată și posibilele soluții.

În cadrul capitolului 4 integrarea filtrelor de imagine pentru a accelera procesarea imaginilor. Prin selectarea și aplicarea atentă a filtrelor, timpul de procesare al imaginilor poate fi redus semnificativ. Experimentele efectuate au evidențiat eficacitatea diferitelor tehnici de filtrare în îmbunătățirea vitezei de procesare cu ~33% fără a compromite acuratețea, prin obținerea unei variații de ~2% pentru Pragul Adaptiv Gaussian.

Capitolul 5 investighează utilizarea ChatGPT pentru accelerarea procesului de dezvoltare a modelelor, evidențiind reducerea semnificativă a timpului pentru sarcinile de dezvoltare și depanare. În mediul nesupervizat, modelele generate au avut o calitate inferioară, însă în abordarea supervizată s-a obținut o acuratețe de 60.08%, cu o arhitectură validată în câteva minute. Aceasta oferă cercetătorului un punct de plecare solid pentru îmbunătățiri ulterioare. Capitolul subliniază importanța selecției cuvintelor și expertiza dezvoltatorilor în echilibrarea vitezei de dezvoltare și a acurateței modelului.

În capitolul 6, autorul propune un studiu în care se implementează un nou set de date creat prin combinarea DailyDialog și FER2013, împreună cu un model multimodal capabil să obțină o acuratețe de 70,19% pe baza de date unică, numit ImaText. Setul de date ImaText constă în combinarea textului cu imagini și conține 25.780 de înregistrări distribuite pe șase emoții. Distribuția finală include 4.865 de emoții de „furios”, 555 de „dezgust”, 1.255 de „frică”, 8.910 de „fericit”, 6.190 de „trist” și 4.005 de „surpriză”. Folosind acest nou set de date, diversele experimente au condus la obținerea unei acurateți de 70,19%. Conform cunoștințelor autorului, ImaText este primul set de date multimodal creat special pentru analiza sentimentelor care include doar text și imagini.

7.2 Contribuţii originale

Printre contribuțiile originale în această lucrare amintim:

1. Crearea unor analize ce descriu aspectele psihologice ale FER și se oferă o descriere a seturilor de date și al algoritmilor care fac posibile rețelele neuronale. Se realizează un studiu al literaturii de specialitate în recunoașterea emoțiilor faciale, detaliind aspectele principale ale cercetărilor, pentru a evidenția noutatea, conceptele și strategiile conexe care fac ca recunoașterea să atingă o precizie bună. De asemenea, sunt evidențiate provocările legate de învățarea automată și posibilele soluții. Aceste provocări oferă o perspectivă asupra direcțiilor posibile de urmat pentru a dezvolta sisteme FER mai bune.

2. Prin selectarea și aplicarea atentă a filtrelor, timpul de procesare al imaginilor poate fi redus semnificativ. Se accelerează procesul de antrenare, în timp ce experimentele efectuate au evidențiat eficacitatea diferitelor tehnici de filtrare în îmbunătățirea vitezei de procesare cu ~33% fără a compromite acuratețea, prin obținerea unei variații de ~2% a acurateței pentru Pragul Adaptiv Gaussian. Această cercetare subliniază beneficiile potențiale ale integrării filtrelor în fluxurile de lucru pentru procesarea imaginilor.

3. Folosirea ChatGPT care reduce semnificativ timpul necesar pentru sarcinile de dezvoltare și de depanare, beneficiind în special la începutul dezvoltării și oferind sugestii pentru remedierea erorilor. În mediul nesupervizat, acesta a generat mai puține modele și de calitate inferioară cu mai multe erori. Pentru abordarea supervizată, rezultatele experimentale subliniază eficacitatea în crearea rapidă a modelelor performante cu un minim de timp investit, obținându-se o soluție cu o acuratețe de 60.08% și o arhitectură validată în câteva minute. De aici, cercetătorul dispune de un punct de plecare bun pentru a îmbunătăți ulterior modelul și a-l ajusta pentru aplicația necesară.

Rezultatele arată importanța selecției cuvintelor și expertiza dezvoltatorilor în atingerea unui echilibru între viteza de dezvoltare și acuratețe.

4. Introducerea unui set de date multimodal numit ImaText. Setul de date constă în combinarea textului cu imagini și conține 25.780 de înregistrări distribuite pe șase emoții. Distribuția finală include 4.865 de emoții de „furios”, 555 de „dezgust”, 1.255 de „frică”, 8.910 de „fericit”, 6.190 de „trist” și 4.005 de „surpriză”. Conform cunoștințelor autorului, ImaText este primul set de date multimodal creat special pentru analiza sentimentelor care include doar text și imagini.

5. Antrenarea unor rețele multimodale pentru a putea determina emoțiile din ImaText. Diversele experimente (folosind configurații diferite de straturi, optimizatori, parametri etc.) au condus la obținerea unei acurateți de 70,19%.

7.3 Lista lucrărilor originale

Această listă cuprinde numai lucrările publicate/communicate la care doctorandul este autor sau co-autor. La acestea se adaugă şi rapoartele de cercetare din programul de doctorat şi contractele la care doctorandul a lucrat. Toate aceste lucrări se regăsesc şi la Bibliografie. Toate lucrările menţionate au un conţinut legat de tematica tezei de doctorat.

1. **Radu-Daniel Bolcaș** and Diana Dranga, "*Facial Emotions Recognition in Machine Learning*," Electrotehnică, Electronică, Automatică (EEA) Journal, vol. 69, no. 4, pp. 87-94, DOI:10.46904/eea.21.69.4.1108010, 2021. **Articol indexat Scopus.**
2. Diana Dranga and **Radu-Daniel Bolcaş**, "*Artificial Intelligence Enhancements in the field of Functional Verification*," Electrotehnică, Electronică, Automatică (EEA) Journal, vol. 69, no. 4, pp. 87-94, DOI:10.46904/eea.21.69.4.1108011, 2021. **Articol indexat Scopus.**
3. **Radu-Daniel Bolcaş**, Mihai Ciuc, and Eduard Popovici, "*Enhancing Training Efficiency in Facial Emotion Recognition*," 2023 31st Telecommunications Forum (TELFOR), pp. 1-4, DOI: 10.1109/TELFOR59449.2023.103726002023, 2023. **Articol indexat IEEExplore**.
4. **Radu-Daniel Bolcaș,** "*Generating FER models using ChatGPT*," Romanian Journal of Information Technology and Automatic Control (RRIA), vol. 34, no. 2, pp. 85-96, 2024. **WOS:001253386000007**. **Articol de revistă ISI**.
5. **Radu-Daniel Bolcaș**, Mihai Ciuc, and Eduard-Cristian Popovici, "*Ensemble models for multimodal sentiment*," U.P.B. Sci. Bull., Series C, vol. 86, no. tbd, p. tbd, 2024. **Articol de revistă ISI** – **acceptat pentru publicare.**

7.4 Perspective de dezvoltare ulterioară

Cercetarea explorează impactul filtrelor de imagine în reducerea timpului de antrenare, subliniind importanța optimizării parametrilor pentru a echilibra viteza și acuratețea. Tehnologia avansată solicită metode eficiente de procesare a imaginilor, iar această cercetare deschide calea pentru utilizarea filtrelor avansate și optimizarea acestora în scenarii reale.

Utilizarea ChatGPT și LLM-urilor pentru validarea rapidă a arhitecturilor promițătoare sugerează potențialul unor investigații viitoare în alte domenii și seturi de date, deși în mediul nesupravegheat modelele generate au fost inferioare. Totuși, modelul bazat pe K-Means a funcționat bine și va fi investigat în continuare.

Modelul multimodal prezentat obține o acuratețe de 70,19%, dar întâmpină dificultăți în recunoașterea anumitor emoții, din cauza unui dezechilibru de clase. Se propune augmentarea datelor sau integrarea seturilor de date suplimentare pentru a rezolva aceste probleme. Integrarea filtrelor în modelul ImaText este, de asemenea, sugerată pentru a crește performanța și a reduce timpul de antrenare.

Bibliografie

[1] Radu-Daniel Bolcaș and Diana Dranga, "Facial Emotions Recognition in Machine Learning," *Electrotehnică, Electronică, Automatică (EEA) Journal*, vol. 69, no. 4, pp. 87-94, DOI:10.46904/eea.21.69.4.1108010, 2021.

[2] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of personality and social psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[3] C. Alberti, K. Lee, and M. Collins, "A bert baseline for the natural questions," *arXiv preprint arXiv:1901.08634*, 2019.

[4] OpenAI. (2024) ChatGPT, last acessed in April 2024. [Online]. https://openai.com/blog/chatgpt

[5] P. Ekman and W. V. Friesen, "Facial Action Coding System," *Consulting Psychologists Press, Stanford University, Palo Alto*, 1977.

[6] Wafa Mellouka and Wahida Handouzi, "Facial emotion recognition using deep learning: review and insights," *Procedia Computer Science*, vol. 175, pp. 689-694, 2020, DOI: 10.1016/j.procs.2020.07.101.

[7] I. J. Goodfellow et al., "Challenges in Representation Learning: A Report on Three Machine Learning Contests," *Neural Information Processing*, pp. p. 117-124, DOI: 10.1007/978-3-642-42051-1_16., 2013.

[8] Z. Rzayeva and E. Alasgarov, "Facial Emotion Recognition using Convolutional Neural Networks," *IEEE 13th International Conference on Application of Information and Communication Technologies (AICT)*, pp. 1-5, DOI: 10.1109/AICT47866.2019.8981757, 2019.

[9] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint, DOI: 10.48550/arXiv.1409.1556*, 2014.

[10] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Emotion recognition using facial expressions," *International Conference on Computational Science*, 2017.

[11] H. Yang, U. Ciftci, and L. Yin, "Facial Expression Recognition by De-expression Residue Learning," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2168-2177, DOI: 10.1109/CVPR.2018.00231, 2018.

[12] Diana Dranga and Radu-Daniel Bolcaş, "Artificial Intelligence Enhancements in the field of Functional Verification," *Electrotehnică, Electronică, Automatică (EEA) Journal*, vol. 69, no. 4, pp. 87-94, DOI:10.46904/eea.21.69.4.1108011, 2021.

[13] Radu-Daniel Bolcaş, Mihai Ciuc, and Eduard Popovici, "Enhancing Training Efficiency in Facial Emotion Recognition," *2023 31st Telecommunications Forum (TELFOR)*, pp. 1-4, DOI: 10.1109/TELFOR59449.2023.103726002023, 2023.

[14] Yousif Khaireddin and Zhuofa Chen, "Facial Emotion Recognition: State of the Art Performance on FER2013," *arXiv preprint, DOI: 10.48550/ARXIV.2105.03588*, 2021. [Online]. https://arxiv.org/abs/2105.03588

[15] OpenCV. (2022). [Online]. https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html

[16] R. C. Gonzalez and R. E. Woods, *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, 2008.

[17] Radu-Daniel Bolcaș, "Generating FER models using ChatGPT," *Romanian Journal of Information Technology and Automatic Control (RRIA)*, vol. 34, no. 2, pp. 85-96, 2024.

[18] Radu-Daniel Bolcaș, Mihai Ciuc, and Eduard-Cristian Popovici, "Ensemble models for multimodal sentiment," *U.P.B. Sci. Bull., Series C*, vol. 86, no. tbd - 3, p. tbd, 2024.

[19] Yanran Li et al., "DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset," *Proceedings of The 8th International Joint Conference on Natural Language Processing (IJCNLP 2017)*, pp. 986-995, 2017.

[20] Shen Weizhou, Siyue Wu, Yunyi Yang, and Xiaojun Quan, "Directed Acyclic Graph Network for Conversational Emotion Recognition," *Annual Meeting of the Association for Computational Linguistics*, 2021.

[21] S. Vignesh, M. Savithadevi, M. Sridevi, and R. Sridhar, "A novel facial emotion recognition model using segmentation VGG-19 architecture," *International Journal of Information Technology*, vol. 15, no. DOI: 10.1007/s41870-023-01184-z, pp. 1777–1787, 2023.