**ROMÂNIA**

**MINISTERUL APĂRĂRII NAŢIONALE**

**ACADEMIA TEHNICĂ MILITARĂ**

**„Ferdinand I”**

**FACULTATEA DE SISTEME INFORMATICE ȘI SECURITATE CIBERNETICĂ**

**Specializarea: Calculatoare şi sisteme informatice pentru apărare**

**şi securitate naţională**



**Aplicație pentru extragerea si indexarea conținutului din documente scanate**

COORDONATOR ȘTIINȚIFIC: ABSOLVENT:

**Lector dr. ing. Stelian Spînu Sd.Plt. Tudose Alin-Romeo**

Conţine \_\_\_\_\_\_\_\_\_\_\_ file Inventariat sub nr. \_\_\_\_\_\_

Poziţia din indicator: \_\_\_\_

Termen de păstrare: \_\_\_\_\_

**BUCUREȘTI**

**2021**

PAGINĂ ALBĂ

PAGINĂ ALBĂ

PAGINĂ ALBĂ

PAGINĂ ALBĂ

# ABSTRACT

# Rezumat

**Cuprins**

[ABSTRACT 6](#_Toc57369109)

[Rezumat 7](#_Toc57369110)

[LISTĂ ABREVIERI 9](#_Toc57369111)

[LISTĂ FIGURI 10](#_Toc57369112)

[1. INTRODUCERE 11](#_Toc57369113)

[1.1. Importanța și actualitatea problematicii abordate 11](#_Toc57369114)

[1.2. Scopul și obiectivele proiectului 11](#_Toc57369115)

[1.3. Rezultate obținute 12](#_Toc57369116)

[1.4. Structura proiectului 12](#_Toc57369117)

[2. STADIUL CUNOAȘTERII ÎN DOMENIU 13](#_Toc57369118)

[2.1. Fundamente teoretice 13](#_Toc57369119)

[2.1.1. Inteligență artificială 13](#_Toc57369120)

[2.1.2. Machine Learning 13](#_Toc57369121)

[2.1.3. Deep Learning 14](#_Toc57369122)

[2.1.4. Problema proiectului 14](#_Toc57369123)

[2.2. Soluții și abordări similare 16](#_Toc57369124)

[2.2.1. Modelul propus de Alex Graves și Jurgen Schmidhuber, din cadrul Universității Tehnice din Munchen 16](#_Toc57369125)

[2.2.2. Modelul propus de Hanu Priya Indiran, din cadrul Universității de Tehnologie Kamaruguru, India 16](#_Toc57369126)

[2.2.3. Modelul propus de Arik Poznanski și Lior Wolf, din cadrul Universității Tel Aviv, „The Blavatnik School of Computer Science” 17](#_Toc57369127)

# LISTĂ ABREVIERI

* IA – Inteligență Artificială
* ML – Machine Learning
* DL – Deep Learning
* CV – Simularea vederii pentru un computer (Computer Vision)
* SP – Procesare de secvențe (Sequence Processing)
* CNN – Rețele Neuronale Convoluționale (Convolutional Neural Networks)
* RNN – Rețele Neuronale Recurente (Recurrent Neural Networks)
* MDRNN – Rețele Neuronale Recurente Multi-Dimensionale (Multi-Dimensional Recurrent Neural Networks)
* CTC – Clasificare Temporală Conexionistă (Connectionist Temporal Classification)
* LSTM – Long Short Term Memory, strat prezent în arhitecturile de RNN.

# LISTĂ FIGURI

...

CAPITOLUL I

# INTRODUCERE

## Importanța și actualitatea problematicii abordate

În societatea modernă, automatizarea proceselor simple, de zi cu zi, a devenit, încetul cu încetul, un subiect de discuție foarte important. Începând de la micile sisteme de IoT, care preiau și excelează in desfășurarea activităților zilnice, precum încuietorile smart sau termostatele automate, care creează mediul optim, după preferințele fiecărei persoane, pana la întregi linii de asamblare industriale, automatizarea tuturor acestor procese a permis ridicarea standardului de viață pentru întreaga omenire, fie că vorbim despre confortul propriu, fie capacitatea de producție a fabricilor în diferite industrii.

Una din problemele ce încă persistă în ziua de astăzi, este procesarea diferitelor tipuri de documente, în mod automat, și cât mai rapid. Rapiditatea, ușurința stocării, si eficiența nu sunt termeni asociați cu modalitatea tradițională de gestionare, unde una sau mai multe persoane erau însărcinate cu gestiunea arhivelor de documente. În plus, întreținerea unei persoane care să se ocupe de această gestiune poate fi costisitoare pe termen lung. Așadar, nevoia implementării unei soluții eficiente, automate, care să reducă atât costurile, cât și timpul de răspuns la diferite tipuri de cereri este actuală.

## Scopul și obiectivele proiectului

Proiectul își propune realizarea unei aplicații software ce va realiza în mod automat extragerea informațiilor din documente formatate, completate de mână. Această aplicație, primind ca input o imagine scanată a unui document preformatat, va identifica sectiunile de interes, din imagine, și anume, informațiile completate de mână. Odată identificate, aceste secțiuni din imagine vor fi oferite unei mașini cu inteligență artificială ce va identifica textul din imagine la nivel de caracter. Caracterele prezise din imagine vor fi asamblate înapoi in cuvinte, si stocate într-o bază de date nerelațională, ElasticSearch. Pentru ușurința accesării și configurării formatelor documentelor, datele salvate vor putea fi accesate printr-o interfață web.

## Rezultate obținute

## Structura proiectului

CAPITOLUL II

# STADIUL CUNOAȘTERII ÎN DOMENIU

## Fundamente teoretice

În ultimii ani, inteligența artificială a fost un subiect de interes, atât pentru media, cât și pentru noii cercetători în domeniul Științei Calculatoarelor. Inteligența artificială, algoritmi de Machine Learning, chiar si algoritmi de Deep Learning, apar în tot mai multe articole, chiar și în afara publicațiilor științifice. Ne-a fost promis un viitor plin de autoturisme autonome și asistenți virtuali inteligenți, de care ne apropiem, din ce în ce mai rapid.

### Inteligență artificială

Inteligența artificială a luat naștere în anii 1950, când o grupare de cercetători în nou-creatul domeniul al științei calculatoarelor, și-au pus următoarea întrebare: *„Ar putea calculatoarele să fie învățate să gândească?”*. Această întrebare este în dezbatere până în ziua de azi. O definiție pentru domeniul IA, ar putea fi următoarea: *„Efortul de a automatiza procese desfășurate, în mod normal, de oameni.”.* Așadar, IA este un domeniu cu un grad foarte mare de generalizare, ce cuprinde domeniile ML și DL.

### Machine Learning

În Anglia epocii victoriene, Lady Ada Lovelace, era o prietenă și colaboratoare a lui Charles Babbage, inventatorul mașinii analitice. Chiar dacă a fost mult inaintea timpului ei, mașina analitică nu a fost niciodata gândită ca un calculator de uz general, atunci când a fost proiectat, în anii 1830 – 1840, deoarece conceptul uzului general al unui calculator nu era înca inventat. A fost doar un mod de a folosi operații mecanice pentru a automatiza anumite calcule din domeniul analizei matematice, de unde și numele acesteia. Ada Lovelace a făcut atunci remarca, despre mașina analitică: *„Mașina Analitică nu are nicio pretenție sa aducă ceva nou. Ea poate să facă tot ce știm noi sa îi spunem să facă. Scopul ei este să ne asiste în găsirea rezultatelor problemelor cu care suntem familiari.”*

Domeniul Machine Learning are la bază această intrebare: *„Poate un calculator să treacă dincolo de granițele a ceea ce știm noi să îi spunem să facă?”*. Așadar, este introdusă o nouă paradigmă de programare. Dacă în programarea clasică, unui calculator îi erau prezentate reguli și date, acesta trebuia să furnizeze rezultate. În domeniul ML, un calculator primește date și rezultate, iar acesta trebuie să furnizeze regulile de obținere a rezultatelor din acele date. Un sistem ML nu este un sistem programat explicit, ci este un sistem antrenat.

### Deep Learning

Deep Learning-ul este un subdomeniu al domeniului ML, o nouă abordare asupra învățării reprezentărilor din date disponibile, ce pune accent pe straturi succesive de învățare. Cuvântul „Deep” din Deep Learning nu presupune vreun fel de înțelegere în profunzime a problemelor, ci doar face referire la modelele cu straturi succesive de învățare. Adâncimea modelului reprezintă numărul de straturi succesive de învățare pe care le are un model. Modelele actuale de DL adesea cuprind zeci, chiar sute de astfel de straturi succesive, și fiecare este antrenat automat prin prezentarea datelor si a rezultatelor așteptate pentru acestea.

### Problema proiectului

Traducerea automată a imaginilor ce conțin scris de mână în seturi de date ce pot fi recunoscute de un computer este un proces destul de dificil. În cadrul acestei teme apar probleme complicate, precum problema segmentării sau problema dicționarului. Pentru a găsi o soluție pentru această problema, este necesar să îmbinăm două din tehnologiile recente de Deep Learning, și anume Computer Vision (Simularea vederii pentru un computer), și Sequence Processing (Procesarea Secvențelor).

Pentru rezolvarea problemelor de Computer Vision, se folosesc, în majoritatea cazurilor, rețele neuronale convoluționale. Acestea introduc operația de convoluție, în straturile lor. Diferența fundamentală între un strat „Dense„ și un strat convoluțional este următoarea: Un strat dens învață trăsaturi reprezentative în întreg spațiul său de intrare, in timp ce un strat convoluțional învață aceste trăsaturi local, în cazul imaginilor, in mici ferestre bidimensionale, de obicei de dimensiuni 3 x 3 sau 5 x 5.

Această caracteristică conferă unui CNN două proprietăți interesante:

1. Trăsăturile învățate de acestea sunt invariante la translatări. După ce a fost învățat un o trăsătura în colțul din dreapta, jos al unei imagini, acesta poate fi recunoscut în orice altă parte a imaginii, de exemplu în stanga, sus.
2. Pot învăța ierarhii spațiale ale trăsăturilor. Un prim strat convolutional va învăța trăsături mici, locale, precum colțuri, sau muchii, un al doilea strat va învăța trăsături mai generale, alcătuite din trăsăturile învățate de primul strat, și așa mai departe. Acest lucru permite CNN-urilor să învețe trăsături din ce în ce mai complexe și abstracte concepte visuale.

Pentru rezolvarea problemelor de Sequence Processing, se folosesc rețele neuronale recurente. Rețelele neuronale obișnuite nu au memorie. Fiecare intrare este procesată independent de celelalte, fără vreo legatură între stările fiecăreia. Dacă vrem să procesăm secvențe, sau date dependente de timp, va trebui să procesăm toată secvența deodată. Comparativ, un RNN, procesează informația prezentă într-o secvență bucată cu bucată, în timp ce reține într-o stare informațiile inputurilor din trecut. În mare, un RNN este o rețea neuronală ce conține o bucla interioară, pentru a accesa informațiile anterioare.

## Soluții și abordări similare

### 2.2.1. Modelul propus de Alex Graves și Jurgen Schmidhuber, din cadrul Universității Tehnice din Munchen

Cei doi cercetători propun un model aparte. În loc să trateze problemele CV si SP separat, precum modelele HMM, folosite pentru transcrieri, aceștia, folosindu-se de două descoperiri recente în domeniul rețelelor neuronale, RNN multidimensionale, și CTC, introduc un model antrenat offline pentru recunoaștere a scrisului de mână, care acceptă la intrare informație sub forma de pixeli de imagine.

Pentru MDRNN se folosesc straturi LSTM. Un strat LSTM este format din celule de memorie conectate in mod recurent, ale caror activări sunt controlate de 3 porți: poarta de intrare, poarta de uitare, și poarta de ieșire. Aceste porți permit celulelor de memorie să rețina, să șteargă și să acceseze informații reținute anterior, obținând recunoașteri dependente de un context destul de larg.

Forma standard a straturilor LSTM, este unidimensionala, deoarece fiecare celulă de memorie conține o singură conexiune recurentă, controlată de o singură poartă de uitare. Totuși, aceste straturi se pot extinde la n dimensiuni, folosind n conexiuni recurente, una pentru fiecare stare anterioară de-a lungul fiecărei dimensiuni, cu n porți de uitare.

CTC este un strat de ieșire creat special pentru probleme de procesare de secvențe folosind RNN-uri. Spre deosebire de alte straturi de ieșire, stratul CTC nu necesită date de antrenare pre-segmentate, și nici post-procesarea ieșirilor, pentru a obține datele recunoscute. Acesta antrenează direct rețeaua să estimeze probabilitățile conditionale ale etichetelor posibile prezente in secvențele de intrare.

Acest model nu are nevoie de niciun fel de preprocesare specifică fiecarui alfabet, deci poate fi folosit neschimbat pentru orice limbă. Dovezile generalității si a eficienței acestuia sunt asigurate de datele puse la dispoziție de o competiție recentă de recunoaștere a caracterelor arabice, unde a obtinut o acuratețe de 91.4%, (câștigătorul concursului a obținut 87.2%), chiar dacă niciunul din autori nu cunosc alfabetul arabic.

### 2.2.2. Modelul propus de Hanu Priya Indiran, din cadrul Universității de Tehnologie Kamaruguru, India

Proiectul lui Indiran dorește să clasifice un cuvânt individual, astfel încât textul scris de mână să poată fi translatat într-o forma digitală, și demonstrează folosirea rețelelor neuronale pentru a proiecta un sistem de recunoaștere a caracterelor din alfabetul englez. Acest sistem preia la intrare imaginile binarizate cu literele ce trebuie să fie recunoscute. Imaginile sunt introduse într-un sistem de extragere de trăsături, al cărui ieșire este luat ca intrare în rețeaua neuronală. Indiran abordează doua metode pentru rezolvare: clasificarea cuvintelor în mod direct, dar și segmentarea acestora și recunoașterea la nivel de caracter.

Pentru prima metodă, diferite arhitecturi de CNN sunt folosite pentru a antrena un model cu acuratețe ridicată, pentru costuri computaționale relativ mici. Dezavantajul acestei metode este însă faptul ca este dependentă de un dicționar de cuvinte inițial, pentru a putea clasifica intrările.

Pentru cea de-a doua metodă, un RNN cu straturi LSTM, sunt folosite, împreună cu convoluția, pentru a crea delimitări pentru fiecare caracter, în imaginile de intrare. După ce au fost extrase imaginile pentru fiecare caracter din imagine, acestea sunt date către un CNN pentru clasificare. În urma clasificării, se reconstruiește fiecare cuvânt, în funcție de rezultatele obținute. Rezultatele obținute vorbesc de la sine, obținându-se o eroare de 1% în recunoașterea cuvintelor. Dacă se folosesc straturi de dimensiune foarte mare pentru CNN, eroarea de recunoaștere poate scădea pana la 0.2%.

### 2.2.3. Modelul propus de Arik Poznanski și Lior Wolf, din cadrul Universității Tel Aviv, „The Blavatnik School of Computer Science”

Proiectul celor doi profesori adopta o metoda relativ aparte. Aceștia folosesc un CNN pentru a estima, plecând de la imaginea unui cuvânt scris de mână, frecvențele grupărilor de n caractere ce apar în cuvânt. Frecvențele monogramelor, digramelor și trigramelor, sunt folosite pentru a asocia profilul estimat al cuvântului de recunoscut, cu profilul real al cuvintelor dintr-un dicționar de dimensiuni mari. Din nou, metoda cu dicționarul nu este o metodă facilă, pentru tema proiectului, din cauza necesității recunoașterii datelor care nu se afla neapărat într-un dicționar, precum nume, prenume, adrese e-mail, etc..

Modelul este bazat pe o arhitectura de tip VGG, formată din straturi convolutionale mici, de dimensiune (3 x 3). Rețeaua are un total de 12 straturi, 9 straturi convoluționale, si 3 straturi dense. Performanțele acestui model sunt destul de ridicate, obținându-se o acuratețe de aproximativ 95%, pe mai multe seturi de date, printre care se numără și seturile IAM, SVT și RIMES.