

## ALIMENTAREA ȘI EVALUAREA SISTEMELOR BAZATE PE ÎNVĂȚAREA PROFUNDĂ UTILIZÂND ENERGIE VERDE

**Teză de doctorat – Rezumat**

pentru obținerea titlului științific de doctor la

Universitatea Politehnică Timișoara

în domeniul de doctorat Calculatoare și tehnologia informației

**autor ing. Sorin Liviu JURJ**

conducător științific Prof.univ.em.dr.ing. Mircea VLĂDUȚIU

luna Septembrie anul 2020

În ultimii ani, progresele în domeniul inteligenței artificiale (AI), în special în ceea ce privește algoritmi Deep Learning (DL) [1], au crescut într-un ritm rapid și vor continua această tendință pentru anii următori. De la implementări hardware la software, studii de cercetare active sunt efectuate în diferite industrii pentru a integra acești algoritmi inspirați de creier în fiecare aspect al vieții noastre. Cu toate acestea, datorită faptului că acești algoritmi necesită o cantitate foarte mare de timp, energie, date și putere de procesare, impactul lor asupra mediului înconjurător este o problemă definitorie [2]. Pentru a rezolva această problemă, având în vedere recente eforturi "Green AI" [3] care se concentrează pe eficiența energetică a sistemelor AI, propunem patru metrici pentru evaluarea performanței modelelor și sistemelor DL bazate nu numai pe acuratețea lor [4, 5], ci și pe consumul de energie și pe costuri.

În această teză de doctorat, am dezvoltat și implementat metode de rezolvare a problemelor menționate mai sus prin implementarea mai întâi a unor aplicații DL diverse care rezolvă diferite probleme legate de fraudă [6, 7] și securitate [8]. Apoi, deoarece am observat că un sistem bazat pe DL care clasifică imagini în timp real [8] consumă mai multă energie decât omologii lor (care nu clasifică imagini în timp real), am decis nu numai să rulăm aceeași implementare pe o platformă care consumă de 5 ori mai puțină energie, dar am vrut, de asemenea, să nu plătim pentru acest consum de energie [9]. Am realizat acest lucru luând în considerare utilizarea energiei verzi [10-12], mai exact, prin construirea unui nou tracker solar cu axă dublă care se bazează pe principiul umbrei (Cast-Shadow) [13] și care a fost ulterior modificat cu costuri minime [9]. Am demonstrat în [9] că tracker-ul nostru solar este eficient și, din câte știm, pentru prima dată în literatură, că este posibil să se utilizeze complet energia solară pentru alimentarea unui sistem bazat pe DL atunci când rulează inferență în timp real.

Pentru a fi conștienți de eventualele defecțiuni ale dispozitivului nostru de urmărire solară cu axă dublă, am investigat, de asemenea, posibilitățile de a-l testa la nivel software și hardware. Pentru aceasta, am implementat o tehnică de testare White-box în [14] și o arhitectură online de autotestare încorporată în sistem (OBIST) în [15], ambele obținând o acoperire ridicată a defectelor. Este important să menționăm că, din câte știm, testarea unui echipament de urmărire solară nu a mai fost făcută niciodată în literatură.

După cum am menționat mai devreme, pentru că am reușit să utilizăm energia verde (solară) pentru alimentarea unui sistem bazat pe DL [9] și pentru că dorim să încurajăm generațiile viitoare de cercetători să i-a în considerare impactul pe care proiectul lor DL îl poate avea asupra mediului înconjurător, am propus patru noi metrici DL [16] care evaluează performanța modelelor și sistemelor DL atât pentru inferență, cât și pentru antrenare, luând în considerare nu numai acuratețea, ci și consumul de energie și costul, dovedindu-se a fi metrici mai valoroase în comparație cu cele existente găsite în literatură, care nu iau în calcul impactul

negativ asupra naturii. De asemenea, am creat o aplicație Computer Vision [17] care încorporează cele patru metrice propuse și care oferă o modalitate ușoară de a calcula și evalua performanța sistemelor bazate pe DL. În plus, aplicația constă din mai multe caracteristici care utilizează inferența DL pentru a accelera sarcinile legate de colectarea [18], curățarea și etichetarea datelor, depășind soluțiile existente cu o marjă mare.

În plus, pentru a oferi studenților de inginerie o șansă de a avea o abordare practică pentru testarea PCB-urilor, am implementat un dispozitiv de testare PCB low-cost și portabil sub forma unui testor de circuite integrate echipat cu sondă mobilă (FPICT) [19] care are o acoperire ridicată a defectelor și un cost foarte scăzut.

În cele din urmă, pentru a crește performanța de trecere a unui Secure Hash Algorithm (SHA)-256, am implementat diferite tehnici pentru a accelera generarea de hash în hardware [20].

Structura tezei de doctorat cuprinde un capitol introductiv, un capitol cu context teoretic, cinci capitole dedicate prezentării cercetărilor realizate și a rezultatelor obținute, un capitol final dedicat concluziilor, contribuțiilor personale și direcțiilor viitoare precum și o listă bibliografică (cu 225 titluri consultate și citate). Teza de doctorat se întinde pe 192 de pagini, demersul de cercetare fiind susținut grafic și sintetic de 91 de figuri și 39 de tabele.

**Capitolul 1, "Introducere"** este dedicat analizei impactului tehnologic și asupra mediului al DL. Obiectivele tezei de doctorat sunt, de asemenea, revizuite începând cu motivația de punere în aplicare a diferitelor aplicații DL legate de fraudă și securitate; de construire, testare și implementare a unui tracker solar cu axă dublă, care este mai târziu utilizat pentru a alimenta un sistem în timp real pe bază de DL folosind 100% energie solară; de propunere a celor patru metrice cât și a aplicației Computer Vision care ajută la calcularea metricilor propuse; de propunere a dispozitivului mobil ieftin de testare a plăcilor de circuite imprimate; de tehnicile de accelerare hardware a unui SHA-256.

**Capitolul 2, "Context teoretic"** prezintă contextul teoretic pentru o mai bună înțelegere a lucrărilor de cercetare care cuprind această teză de doctorat și care sunt prezentate începând cu capitolul 3. Prezentăm unele dintre arhitecturile de rețele neuronale, framework-uri și analiza seturilor de date pentru diferite aplicații DL.

De asemenea, acoperim o secțiune legată de testarea hardware și software care descrie diferite metode de testare off-line și on-line pe care le-am utilizat. În cele din urmă, lucrările conexe cu privire la aplicațiile noastre DL, la testarea hardware și software, la platforma hardware cu consum redus, metrice, colectarea datelor și etichetarea lor, precum și în ceea ce privește implementările hardware ale SHA-256 sunt, de asemenea, prezentate.

**Capitolul 3, "Diferite aplicații bazate pe Deep Learning pentru detectarea fraudei și creșterea securității"** prezintă 3 aplicații diferite de clasificare a imaginilor DL:

a) o metodă nouă în detectarea fraudei într-un Supermarket folosind o aplicație smartphone care utilizează un algoritm OCR compus din tehnici de procesare a imaginilor și rețele neuronale convoluționale (CNNs). Metoda propusă detectează cu succes prețurile de la etichetele de preț ale produsului, precum și prețurile de pe chitanțe cu mare precizie. În plus, modelele CNN propuse depășesc alți algoritmi OCR open-source cunoscuți în ceea ce privește acuratețea pe imagini cu prețurile trunchiate ale unui produs și de pe chitanță, care conțin chiar și „zgomot”;

b) o metodă nouă de identificare a motivelor tradiționale românești găsite pe 4 categorii (haine, ceramică, covoare și ouă pictate) folosind CNN-uri. Am implementat, de asemenea, un sistem care poate detecta și identifica aceste motive învățate printr-o cameră web cu o precizie ridicată și timp redus de procesare;

c) o metodă nouă de identificare a animalelor care aparțin celor mai cunoscute 34 de specii găsite în zonele domestice ale Europei. Am implementat un sistem care poate identifica aceste specii în imagini, clipuri video sau printr-o cameră web cât și genera 2 seturi de date noi

în timp real, unul care conține informații textuale despre animalul prezent în fața camerei web, și unul care conține imagini ale speciilor de animale identificate. Metoda noastră are mai multe avantaje în comparație cu alte lucrări conexe.

Conținutul acestui capitol se bazează în principal pe lucrările noastre în [6-8].

**Capitolul 4, "Alimentarea unui sistem Deep Learning în timp real folosind energia solară"** prezintă construcția, testarea și implementarea unui tracker solar cu axă dublă pentru a alimenta un sistem bazat pe DL în timp real. Mai exact:

a) o tehnică de testare în verificarea codului software care rulează pe un microcontroler Arduino UNO care optimizează poziția unui dispozitiv de urmărire solară prin recurgerea la elemente noi precum comutatoari de limitare și elemente de blocare. Metoda noastră software a contribuit la o reducere semnificativă a consumului de energie și a dovedit eficiența versiunilor automate ale panourilor solare față de cele statice;

b) o metodă White-Box de testare aplicată pe un tracker solar care testează codul software care rulează pe un modul Wi-Fi, Nodemcu Lua ESP8266, și care dovedește că este eficientă din punct de vedere al acoperirii erorilor cât și a costurilor. În plus, dobândim capacitatea de a controla direct mișcările motorului pas cu pas ale trackerului solar autonom într-un mod wireless;

c) o tehnică de testare hardware care utilizează un OBIST și care intervine în testarea echipamentelor electrice ale unui dispozitiv de urmărire solară pentru eventuale defecțiuni hardware, cu scopul de a minimiza costurile de funcționare și care este eficient în ceea ce privește acoperirea de erori hardware;

d) o metodă nouă în alimentarea unui sistem bazat pe DL în timp real, folosind 100% energie verde, utilizând o platformă încorporată Nvidia Jetson TX2 și un tracker solar îmbunătățit cu axă dublă, care a fost conectat la un lanț de două invertoare, un acumulator și un regulator de încărcare solară. Modificările noastre de implementare a software-ului ajută la detectarea utilizării optime a memoriei GPU și a cadrelor pe secundă (fps) pentru a rula modelele noastre DL fără nici un risc de a avea erori de tip "out of memory", iar împreună cu o metodă software de detectare a mișcării, reușim să reducem și consumul de energie al întregului sistem bazat pe DL.

Conținutul acestui capitol se bazează în principal pe lucrările noastre în [9, 13-15].

**Capitolul 5, "Metrici propuse pentru Deep Learning"** prezintă metricile propuse pentru DL, precum și o aplicație Computer Vision cu mai multe capacități încorporate orientate spre Data Science. Mai exact:

a) cele patru metrici noi, precizie per consum (APC), precizia per cost de energie (APEC), timpul până la cel mai apropiat APC (TTCAPC) și timpul până la cel mai apropiat APEC (TTCAPEC) pentru evaluarea performanței modelelor și sistemelor DL nu numai în ceea ce privește acuratețea, ci și consumul de energie și costul lor, arătând că sistemele bazate pe DL alimentate cu energie verde sunt evaluate ca fiind mult mai performante în comparație cu cele existente care încă utilizează o rețea tradițională ca sursă de energie;

b) o aplicație cu o interfață ușor de utilizat care rezolvă multe probleme legate de pregătirea datelor necesare antrenării unui model DL și care oferă o modalitate ușoară de a evalua performanța de sisteme bazate pe DL folosind calculatoare pentru metricile noastre APC, APEC, TTCAPC și TTCAPEC.

Conținutul acestui capitol se bazează în principal pe lucrările noastre din [16, 17].

**Capitolul 6, „Testor de circuite integrate echipat cu sondă mobilă, cu aplicabilitate în Educație și Ingineria Testării”** prezintă un FPICT accesibil, portabil și ușor de utilizat, care are scopuri educaționale în domeniul ingineriei de testare, în principal pentru testarea PCB-urilor de dimensiuni mai mici, cum ar fi Arduino Uno, fără a avea nevoie de senzori. FPICT-ul propus poate fi ușor conectat la orice platformă de calcul care are un port USB și fișierele sale de configurare scrise în limbajul de programare C pot fi modificate cu ușurință, oferind studenților acces ușor pentru a studia și experimenta cu funcționarea interioară a unui Flying

Probe Tester (FPT) atunci când testează o placă PCB reală.

Conținutul acestui capitol se bazează în principal pe activitatea noastră în [19].

**Capitolul 7, "Soluții tehnologice pentru îmbunătățirea trecerii ale unui Secure Hash Algorithm-256"** prezintă mai multe tehnici de accelerare pentru îmbunătățirea trecerii unei implementari hardware SHA-256. În primul rând, tehnica de accelerare a trecerii elimină un tact de ceas utilizat pentru actualizarea valorii hash și permite livrarea unei treceri mai mari. De asemenea, calea critică a unei structuri Carry-Save-Adder (CSA) este considerabil redusă prin utilizarea unui Kogge-Stone adder rapid pe 32 de biți. Cu cea de-a doua tehnică, am evaluat structuri alternative de completare multioperand și am implementat Carry-Propagate-Adder-urile adders-ului multioperand într-un mod fuzionat pentru a accelera generarea funcțiilor rotunde. Abordarea bazată pe sinteză pentru aranjarea ordinii operanzilor (delay balancing improvement) în arborele CSA reduce în continuare calea critică.

Conținutul acestui capitol se bazează în principal pe activitatea noastră în [20].

**Capitolul 8, "Concluzii și lucrări viitoare"** prezintă concluziile acestei teze de doctorat precum și lucrările viitoare. În plus, prezintă cele 11 lucrări publicate în timpul studiilor mele de doctorat, toate indexate ISI, din care 3 sunt, de asemenea, și capitole de carte Springer.

### **Bibliografie selectivă:**

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning" Nature, vol. 521, no. 7553, pp. 436–444, 2015

[2] Strubell, E., Ganesh, A., McCallum, A.: Energy and policy considerations for deep learning in NLP. In: arXiv:1906.02243, (2019)

[3] Schwartz, R., Dodge, J., Smith, N.A., Etzioni, O.: Green AI. In: arXiv:1907.10597v3, August, (2019).

[4] Mattson, P., et al.: MLPerf Training Benchmark. In: arXiv:1910.01500v2, October (2019).

[5] Reddi, V.J., et al.: MLPerf Inference Benchmark. In: arXiv:1911.02549, November (2019).

[6] Sorin Liviu Jurj, Flavius Opritoiu, Mircea Vladutiu „Identification of Traditional Motifs using Convolutional Neural Networks”, 2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME), Iasi, Romania, pp. 191-196, 2018

[7] Sorin Liviu Jurj, Allen-Jasmin Farcas, Flavius Opritoiu, Mircea Vladutiu „Mobile Application for Receipt Fraud Detection Based on Optical Character Recognition”, Proc. SPIE 11433, Twelfth International Conference on Machine Vision (ICMV 2019), 1143313 (31 January 2020);

[8] Jurj, S.L., Opritoiu, F., Vladutiu, M.: Real-time identification of animals found in domestic areas of Europe. In: Proc. SPIE 11433, Twelfth International Conference on Machine Vision (ICMV 2019), 1143313 (31 January 2020). doi: 10.1117/12.2556376.

[9] Jurj S.L., Rotar R., Opritoiu F., Vladutiu M. (2020) Efficient Implementation of a Self-sufficient Solar-Powered Real-Time Deep Learning-Based System. In: Iliadis L., Angelov P., Jayne C., Pimenidis E. (eds) Proceedings of the 21st EANN (Engineering Applications of Neural Networks) 2020 Conference. EANN 2020. Proceedings of the International Neural Networks Society, vol 2. Springer, Cham, pp. 99-118, doi: 10.1007/978-3-030-48791-1\_7.

- [10] Ram M., et al. Global Energy System based on 100% Renewable Energy – Power, Heat, Transport and Desalination Sectors. Study by Lappeenranta University of Technology and Energy Watch Group, Lappeenranta, Berlin, March 2019. [Online]. Available: [http://energywatchgroup.org/wp-content/uploads/EWG\\_LUT\\_100RE\\_All\\_Sectors\\_Global\\_Report\\_2019.pdf](http://energywatchgroup.org/wp-content/uploads/EWG_LUT_100RE_All_Sectors_Global_Report_2019.pdf)
- [11] Yan, J., Yang, Y., Elia Campana, P., He, J.: City-level analysis of subsidy-free solar photovoltaic electricity price, profits and grid parity in China. *Nat. Energy*(4), 709–717, (2019).
- [12] Rolnick, D., et al.: Tackling Climate Change with Machine Learning. In: arXiv:1906.05433v2, November (2019).
- [13] Raul Rotar, Sorin Liviu Jurj, Flavius Opritoiu, Mircea Vladutiu, “Position Optimization Method for a Solar Tracking Device Using the Cast-Shadow Principle”, 2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME), pp. 61-70, (2018)
- [14] Sorin Liviu Jurj, Raul Rotar, Flavius Opritoiu, Mircea Vladutiu, "White-Box Testing Strategy for a Solar Tracking Device using NodeMCU Lua ESP8266 Wi-Fi Network Development Board Module", 2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME), pp. 53-60, 2018
- [15] Sorin Liviu Jurj, Raul Rotar, Flavius Opritoiu, Mircea Vladutiu „Online Built-In Self-Test Architecture for Automated Testing of a Solar Tracking Equipment”, 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Madrid, Spain, 2020, pp. 1-7, doi: 10.1109/EEEIC/ICPSEurope49358.2020.9160850.
- [16] Jurj, S.L., Opritoiu, F., Vladutiu, M.: Environmentally-Friendly Metrics for Evaluating the Performance of Deep Learning Models and Systems. In *Neural Information Processing, Lecture Notes in Computer Science (LNCS)*, Proceedings of the 27th International Conference on Neural Information Processing (ICONIP 2020), Bangkok, Thailand, (November 2020). To appear.
- [17] Jurj S.L., Opritoiu F., Vladutiu M. (2020) Deep Learning-Based Computer Vision Application with Multiple Built-In Data Science-Oriented Capabilities. In: Iliadis L., Angelov P., Jayne C., Pimenidis E. (eds) *Proceedings of the 21st EANN (Engineering Applications of Neural Networks) 2020 Conference*. EANN 2020. Proceedings of the International Neural Networks Society, vol 2. Springer, Cham, pp. 47-69, doi: 10.1007/978-3-030-48791-1\_4.
- [18] Roh, Y., Heo, G., Whang, S.E.: A Survey on Data Collection for Machine Learning: a Big Data -- AI Integration Perspective. In: arXiv:1811.03402v2, August, (2019)
- [19] Sorin Liviu Jurj, Raul Rotar, Flavius Opritoiu, Mircea Vladutiu „Affordable Flying Probe-Inspired In-Circuit-Tester for Printed Circuit Boards Evaluation with Application in Test Engineering Education”, 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Madrid, Spain, 2020, pp. 1-6, doi: 10.1109/EEEIC/ICPSEurope49358.2020.9160639.
- [20] Flavius Opritoiu, Sorin Liviu Jurj, Mircea Vladutiu „Technological solutions for

throughput improvement of a Secure Hash Algorithm-256 Engine”, 2017 IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME), Constanta, Romania, pp. 159-164, 2017