
Project 1 - Cancer de sân - screening

O echipă de medici povestește:

Cancerul de sân, cea mai comună formă de cancer la nivel mondial, afectează un număr foarte mare de femei. În România, în anul 2020, peste 12 000 de femei au fost depistate ca suferind de această boală. Mai mult, stadiul bolii este foarte avansat la aproximativ 80% din cazuri. În acest context și corelat cu recomandările Uniunii Europene EU Council 2003/878/EC, se dorește dezvoltarea unor programe de screening care să ajute la identificarea persoanelor afectate încă din primele faze ale bolii. Programele de screening pot reduce rata mortalității unei populații cu până la 30%.

Problemă:

Dezvoltarea unui sistem de evaluare și auditare a programelor de screening pentru cancerul de sân, bazat pe recomandările UE și tehnici inteligente de analiză a datelor.

Date:

European Cancer Data available – aggregated

- <https://ecis.jrc.ec.europa.eu/>
- https://ec.europa.eu/health/sites/health/files/major_chronic_diseases/docs/2017_cancerscreening_2ndreportimplementation_en.pdf

Consultant:

Atila

Tasks:

- studiu comparativ al indicatorilor agregati de screening și corelarea lor cu zonele de incidență
 - old guideline <https://www.euref.org/european-guidelines>
 - new guideline <https://bmccancer.biomedcentral.com/articles/10.1186/s12885-020-07289-z>
 - extragerea datelor din baza de date europeană, calcularea indicatorilor (vechi și noi) și compararea lor
 - construirea unui sistem de recomandare de prevenție

Referință:

1. McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafiyan, H., ... & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89-94. <https://www.nature.com/articles/s41586-019-1799-6#Sec8>

Proiect 2 - Cancer de sân – diagnostic

O echipă de medici povestește:

Cancerul de sân, cea mai comună formă de cancer la nivel mondial, afectează un număr foarte mare de femei. În România, în anul 2020, peste 12 000 de femei au fost depistate ca suferind de această boală. Mai mult, stadiul bolii este foarte avansat la aproximativ 80% din cazuri. În acest context și corelat cu recomandările Uniunii Europene EU Council 2003/878/EC, se dorește dezvoltarea unor programe de screening care să ajute la identificarea persoanelor afectate încă din primele faze ale bolii. Programele de screening pot reduce rata mortalității unei populații cu până la 30%.

Problemă:

Dezvoltarea unui sistem de asistență în diagnosticul pentru cancerul de sân, bazat pe recomandările UE și tehnici inteligente de analiză a datelor.

Date:

Mamografii

- MIAS, DDSM, IRMA <https://www.mammoimage.org/databases/>
- Un set nou – menționat aici <https://www.nature.com/articles/s41586-019-1799-6#Sec8>

Consultant:

Cristiana

Tasks:

- studiu comparativ al metodelor de analiză imagistică utile în procesarea mamografiilor
 - normalizarea datelor (dimensiune, regiuni de interes)
 - extragerea de caracteristici din imagini
 - antrenarea și testarea unui clasificator de imagini (sanatoase, benigne, maligne)

Referințe:

1. Abdelhafiz, D., Yang, C., Ammar, R., & Nabavi, S. (2019). Deep convolutional neural networks for mammography: advances, challenges and applications. *BMC bioinformatics*, 20(11), 1-20. <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2823-4#Sec29>
2. Ahmed, L., Iqbal, M. M., Aldabbas, H., Khalid, S., Saleem, Y., & Saeed, S. (2020). Images data practices for semantic segmentation of breast cancer using deep neural network. *Journal of Ambient Intelligence and Humanized Computing*, 1-17. <https://link.springer.com/article/10.1007/s12652-020-01680-1>

Project 3 - Chat bot medical

Un pacient ne povestește:

Am fost azi dimineață la un control medical și doctorul, după o serie de investigații, a început să-mi spună ce ar trebui să fac pentru a mă simți mai bine, ce medicamente să iau, în ce ordine, etc. Dar mi-a zis așa multe deodată încât s-au amestecat toate și e cam haos în mintea mea. Mi-a dat o rețetă, dar e cu scris de doctor...

Un medic rezident ne povestește:

Au trecut deja două săptămâni din rezidențiatul meu. Am avut șansa să intru într-o echipă excelentă de medici profesioniști de la care pot învăța foarte multe. Dar îmi transmit așa multe informații, încât mi-e greu să le rețin și să le asimilez. Nu apuc să notez mai nimic...

Problemă:

Cum pot înțelege limbajul vocal al medicilor? Să îl recunosc și să îmi iau notițe.

Date:

<https://commonvoice.mozilla.org/ro/datasets>

Consultant:

Ioana Alexandra Todericiu (ioana.todericiu@stud.ubbcluj.ro)

Tasks:

Transcrierea mesajelor vocale medicale în text. Pentru că ne adresăm medicilor care vorbesc în limba română trebuie investigat dacă transcrierea se poate realiza direct în limba română sau indirect (printr-o limbă auxiliară precum engleză sau franceză).

Referințe:

1. Speech Representation wav2vec
 - <https://ai.facebook.com/blog/wav2vec-state-of-the-art-speech-recognition-through-self-supervision/>
 - <https://github.com/pytorch/fairseq/tree/master/examples/wav2vec?fbclid=IwAR3GY5uz-Ft2HWfQKUmz1NNueTNhamtvjdVwLdKheahdJZmgCv0kiZuiNM>
2. Google speech to text <https://cloud.google.com/speech-to-text/docs/basics#select-model>
3. Amazon transcribe medical <https://docs.aws.amazon.com/transcribe/latest/dg/transcribe-medical.html>
4. Kodish-Wachs, J., Agassi, E., Kenny III, P., & Overhage, J. M. (2018). A systematic comparison of contemporary automatic speech recognition engines for conversational clinical speech. In *AMIA Annual Symposium Proceedings* (Vol. 2018, p. 683). American Medical Informatics Association. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371385/>
5. Quiroz, J. C., Laranjo, L., Kocaballi, A. B., Berkovsky, S., Rezazadegan, D., & Coiera, E. (2019). Challenges of developing a digital scribe to reduce clinical documentation burden. *NPJ digital medicine*, 2(1), 1-6. <https://www.nature.com/articles/s41746-019-0190-1.pdf?origin=ppub>
6. Schloss, B., & Konam, S. (2020, September). Towards an automated soap note: Classifying utterances from medical conversations. In *Machine Learning for Healthcare Conference* (pp. 610-631). PMLR. <http://proceedings.mlr.press/v126/schloss20a/schloss20a.pdf>

Project 4 - Analiza statistică multivariată a datelor spectrale pentru diagnostic medical

O echipă de fizicieni ne povestește:

"În cadrul studiilor noastre interogăm prin metode spectroscopice probe de biofluide (ser sanguin, salivă, urină...). Spectrele conțin informație structurală la nivel molecular, astfel că putem identifica anumite molecule (metaboliți) a căror detecție poate fi utilă în diagnosticul medical. Subliniem aici că identificarea tipurilor de molecule se obține din forma spectrului (tradus, analiză de semnal). În studiile noastre analizăm, de regulă, 30-100 probe de la pacienți cu o anumită patologie și de asemenea probe de la subiecți sănătoși (control). Apoi, spectrele înregistrate sunt procesate (normalizare, background subtraction...) și analizate prin metode statistice multivariate (supervizate și nesupervizate) pentru a clasifica probele. Cele mai uzuale modele de discriminare folosite de noi sunt PCA, LDA și diferite tipuri de machine learning."

Problemă:

Analiza unui set de date format din pacienți ce au suferit transplant renal (123 pacienți) tratați cu un imunosupresor. O parte din pacienți suferă de intoxicație în urma tratamentului cu imunosupresor. Obiectivul studiului este de a dezvolta o metodă rapidă, relativ simplă și ieftină de detecție a intoxicației pe bază de spectroscopie optică.

Date:

Spectre optice de la ser sanguin și urină + date de spectrometrie de masă (MS) de la probele de ser și urină – unde sunt identificați explicit metaboliții (metabolomică).

Consultant:

Ștefania Iancu (stefania.iancu22@yahoo.ro)
Matyas
Adel

Tasks:

- identificarea metodei optime de preprocesare (feature selection) și clasificare nesupervizată a spectrelor în cele 2 clase (intoxicat versus sanatos) cu metode precum
 - PCA (Principal Component Analysis)
 - LDA (Linear Discriminant Analysis)
 - t-SNE
 - Auto-encoders
- Clasificarea doar a spectrelor optice, clasificarea doar a valorilor MS, clasificarea datelor fuzionate (optice și valori MS)
- corelarea spectrelor optice cu valori din datele MS (metabolice).

Referințe:

1. Moisoiu, V., Ștefancu, A., Gulei, D., Boitor, R., Magdo, L., Raduly, L., ... & Berindan-Neagoe, I. (2019). SERS-based differential diagnosis between multiple solid malignancies: Breast, colorectal, lung, ovarian and oral cancer. *International journal of nanomedicine*, 14, 6165.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6684856/>
2. Wang, C., Xiao, L., Dai, C., Nguyen, A. H., Littlepage, L. E., Schultz, Z. D., & Li, J. (2020). A Statistical Approach of Background Removal and Spectrum Identification for SERS Data. *Scientific reports*, 10(1), 1-11. <https://www.nature.com/articles/s41598-020-58061-z>
3. Morais, C. L., Lima, K. M., Singh, M., & Martin, F. L. (2020). Tutorial: multivariate classification for vibrational spectroscopy in biological samples. *Nature Protocols*, 15(7), 2143-2162.
<https://www.nature.com/articles/s41596-020-0322-8>

Project 5 - Analiza datelor obținute cu microscop hiperspectral pe probe biologice

O altă echipă de fizicieni ne povestește:

Avem un microscop hiperspectral de la Cytoviva (<https://cytoviva.com/online-resources/> - <https://youtu.be/GEj8J9mmq9M>).

Cu ajutorul microscopului obținem serii de date de la probe biologice și/sau nanoparticule metalice sub forma unor spectre în domeniul 400-1000 nm pentru fiecare pixel din imaginea microscopică. Spectrele conțin informații despre emisia/împrăștierea luminii pe proba de interes.

Datele sunt structurate într-un hipercub 4-dimensional (3 dimensiuni spațiale – coordonatele punctului din probă și o dimensiune spectrală), în format [ENVI](#) - softul microscopului este construit pe acest soft geospațial.

Noi folosim aceste date acum numai prin filtrare directă în softul ENVI, pentru a selecta pixeli unde cunoaștem semnalul probei de interes. De exemplu, nanoparticulele de aur au un semnal spectral specific la 600 nm; vom segmenta imaginea microscopică astfel încât să păstrăm doar pixelii care conțin un semnal la valoarea de 600 nm. Dar știm că datele conțin mult mai mult decât atât iar printr-o segmentare a imaginilor, bazată fie pe imaginea microscopică propriu-zisă fie pe spectrul din fiecare punct al imaginii, s-ar putea obține mult mai multe informații.

Problemă:

Familiarizarea cu setul de date și obținerea unei metode de filtrare/procesare a datelor cu posibilitatea de vizualizare/accesare și extragere a informațiilor suplimentare:

Date:

Hipercub 4D - spectre optice pentru fiecare pixel dintr-o imagine despre probe biologice obținute cu microscopul hiperspectral Cytoviva.

Consultant:

Ștefania Iancu (stefania.iancu22@yahoo.ro)

Tasks:

- structurarea datelor hiperspectrale într-un format accesibil.
- corelarea datelor cu imaginea optică/fluorescentă a probei.

Referință:

Imagini/spectre despre țesutul din studiul nostru imATFIB (fig. 7 din articol): <https://www.dovepress.com/assessment-of-gold-coated-iron-oxide-nanoparticles-as-negative-t2-cont-peer-reviewed-article-IJN>

Proiect 6 - Asistent automat în circulația feroviară

O echipă de programatori povestește:

”...ce-ar fi ca pe lângă ajutorul acordat în asistența auto, să dezvoltăm sisteme de decizie utile și în traficul feroviar...să detectăm șinele, semnalele, obstacolele, etc...”

Problemă:

Recunoașterea obiectelor (șine, semnale, stâlpi, etc.) în traficul feroviar. .

Date:

RailSem19

- https://openaccess.thecvf.com/content_CVPRW_2019/papers/WAD/Zendel_RailSem19_A_Dataset_for_Semantic_Rail_Scene_Understanding_CVPRW_2019_paper.pdf
- <https://wilddash.cc/railsem19>

Consultant:

Mircea Suceveanu (mirceasuceveanu@yahoo.com)

Tasks:

- studiu comparativ al metodelor de segmentare semantica a șinelor de tren
- studiu comparativ al metodelor de segmentare semantica a semnalelor pentru tren

Referințe:

1. Gong, X., Chen, X., & Chen, W. (2020). Enhanced Few-shot Learning for Intrusion Detection in Railway Video Surveillance. *arXiv preprint arXiv:2011.04254*. <https://arxiv.org/pdf/2011.04254v1.pdf>
2. SMART project <http://www.smartrail-automation-project.net/index.php/results/deliverables>
3. Ristić-Durrant, D., Haseeb, M. A., Franke, M., Banić, M., Simonović, M., & Stamenković, D. (2020, September). Artificial Intelligence for Obstacle Detection in Railways: Project SMART and Beyond. In *European Dependable Computing Conference* (pp. 44-55). Springer, Cham. https://www.researchgate.net/publication/343985162_Artificial_Intelligence_for_Obstacle_Detection_in_Railways_Project_SMART_and_Beyond