RESEARCH ARTICLE

Artificial intelligence based model for establishing the histopathological diagnostic of the cutaneous basal cell carcinoma

Andrei Calin Dragomir[1], Iuliu Gabriel Cocuz[1]*, Ovidiu Simion Cotoi[2], Leonard Azamfirei[3]

1. George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania
2. Department of Pathophysiology, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania
3. Department of Anesthesia and Intensive Care, George Emil Palade University of Medicine, Pharmacy, Science, and Technology of Targu Mures, Romania

Introduction: Artificial intelligence (AI), a component of computer science, has the ability to process the multitude of medical data existing in the medical system around the world. The goal of our study is to build an AI model, based on Machine Learning, capable of assisting pathologists around the world in the diagnosis of the basal cell carcinoma of the skin. **Material and Method**: Our study is represented by the development of a Mask-RCNN (Mask Region-based Convolutional Neural Network) model, for the detection of cells with typical basal cell carcinoma tumoral changes. A number of 258 digitized histological images were used. The images emerged from Hematoxylin&Eosin stained pathology slides, diagnosed with cutaneous basal cell carcinoma between January 2018 and December 2021, at the Pathology Service of the Mureș County Clinical Hospital. **Results**: All the used images have the unique resolution of 2560x1920 pixels. For the learning process, we divided these images into two datasets: the learning dataset, representing 80% of the total images; and the test dataset, representing 20% of the total images. The AI model was trained using 1000 epochs with a learning rate of 0.00025 and only one classification category: basal cell carcinoma. **Conclusions**: The AI model successfully identified in 85% of the cases the areas with pathological changes present in the input images.

Keywords: histopathology, AI, Mask-RCNN, cutaneous basal cell carcinoma

Introduction

Artificial intelligence (AI) is a complex part of computer science that aims to create intelligent algorithms with the ability to perform functions normally performed by human intelligence. The best-known functions of AI include visual perception, speech recognition, decision-making and translation, tasks that usually require the process of human brain. AI can be considered an interdisciplinary science. The term of „AI" is usually associated with devices such as computers and robots [1-4].

AI systems that meet the needs of the medical field follow a predefined pattern, which can be used in almost all medical fields. For the proper functioning and accuracy of the results and decisions generated by AI systems, it is necessary to create and introduce large databases, and then, based on them, machine learning algorithms are used to generate new information [5,6].

In Pathology, the doctor generates, analyzes and integrates large volumes of data coming from various sources: information from patient observation sheets, data acquired by analyzing histological slides using basic stains (Hematoxylin-Eosin) or special stains, as well as by through immunohistochemistry or molecular biology. In the last two decades, the concept of digital pathology has been constantly developed, the digital era having an impact on histopathological diagnosis. The COVID-19 pandemic has

indicated the need to develop this field, both for diagnosis and for the training of pathologists [7].

Digitization of histopathological slides serves as an excellent source of information through digital morphometry techniques. The complexity and accuracy of the images obtained by digitization is greater than in other medical fields that use imaging due to their large size of the obtained images (a resolution of 100k × 100k is common), the presence of information about the color of the structures observed on the slides, the availability of information on more many scales (eg, ×4, ×20), as well as different levels of sectioning of the paraffin block, all of which realize the premise of using AI in digital pathology. Until now, AI has mainly been used for image-based diagnosis in radiology and cardiology. Its applicability in pathology is a current topic in current research [8-10].

The objective of the study was to build an AI-based model that can be easily used and implemented in the process of computer-aided histopathological diagnosis of cutaneous basal cell carcinoma (BCC), based on digital images obtained by scanning slides in the basic histological staining Hematoxylin- Eosin.

Material and methods
Establishing the database
The database used for training the AI model for the diagnosis of BCC is represented by 258 digitized histological images. The images come from tissue samples from the Pathological Anatomy Service of the Mureş County

* Correspondence to: Iuliu Gabriel Cocuz
E-mail: iuliu.cocuz@umfst.ro

Clinical Hospital. The tissue samples were prepared and histoprocessed according to the standard method, with Hematoxylin-Eosin basic staining performed for histopathological diagnosis.

The study included 208 patients who were diagnosed with BCC between January 2018 and December 2021. The tissue samples came from the Departments of Dermatology, Plastic Surgery and General Surgery of the Mureș County Clinical Hospital. The inclusion criteria of the histopathological preparations in the study were represented by the histopathological diagnosis of cutaneous basal cell carcinoma, respectively by the diagnosis period between January 2018 – December 2021

Data collection and interpretation

The histological slides were scanned and digitized using the Zeiss Axio microscope and the ZenPro 3.2 image acquisition program at a unique resolution of 2560x1920 pixels, with the AI model requiring only one type of resolution for the entire dataset to function properly.

Development of the Machine Learning model

For the learning process of the Machine Learning (ML) model, the data required preprocessing, by polygon marking the areas with pathological changes for each individual histopathological image. For this process we used *labelme*, a graphical image annotation tool. The successful implementation of the program required two components of the learning process. After labeling the histopathological images, the images were divided into two datasets:

- Learning Dataset – consisting of 80% of the images included in the study
- Testing Dataset – consisting of 20% of the images included in the study

The aim of our study was to build an AI model capable of analyzing histological images, by analyzing the pixels of each individual cell, categorizing them and identifying abnormal cells. For this task, in AI, Convolutional Neural Network (CNN) is the gold standard for image classification. However, in this step, our goal is not to classify the input images, but to segment them. CNNs have undergone many changes over time, leading to Region-based Convolutional Neural Network (RCNN) models, then Fast-RCNN, Faster-RCNN, and finally Mask-RCNN. Mask-RCNN models can identify a precise mask in the given image, thus segmenting the object of interest within that image. In this case, the model will learn to segment the images extracted from the histopathology slides to identify the pathology in question.

Images are received as bytes and converted into .jpg images. Figure 2 exemplifies the process that each image goes through from the moment it arrives in the Mask-RCNN model. Each individual histological image will go through the process described previously. The algorithm will later save all the segmented images to a new location. Following this process, the algorithm will return images with the
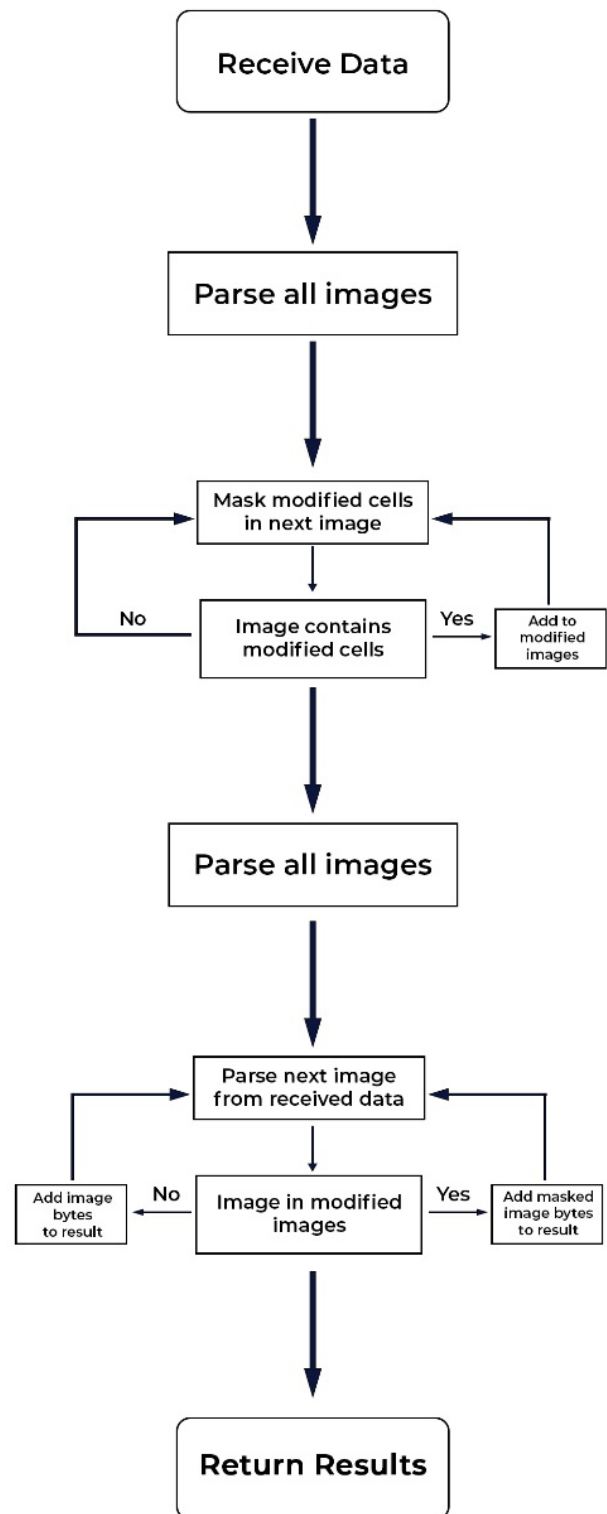

Fig. 1. Schematic illustration of image processing by the AI model

area that has undergone pathological changes, highlighted as masks on the original image. An example can be seen in Figure 3.

Results
Algorithm results

The algorithm results were obtained by combining Intersection Over Union (IOU) and Dice Similarity Coefficient (DSC) values. The IOU was calculated by combining the
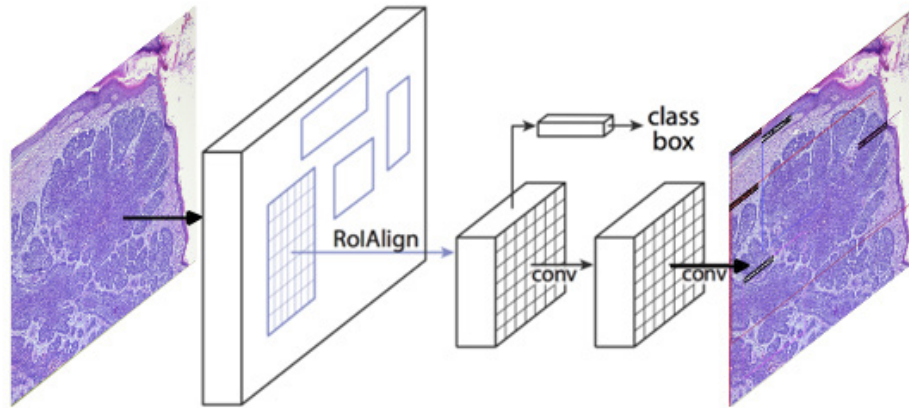
Fig. 2. Schematic illustration of the Mask RCNN model. RolAlign - Region of Interest Align, a layer used for identifying exact spatial location of each pixel for the input image, Class box - a binary classification for each class independently (e.g.: Basal cell carcinoma), Conv - convolution, the process where the algorithm process group of pixels from the input image and transmit the information to the next layer (group of pixels)
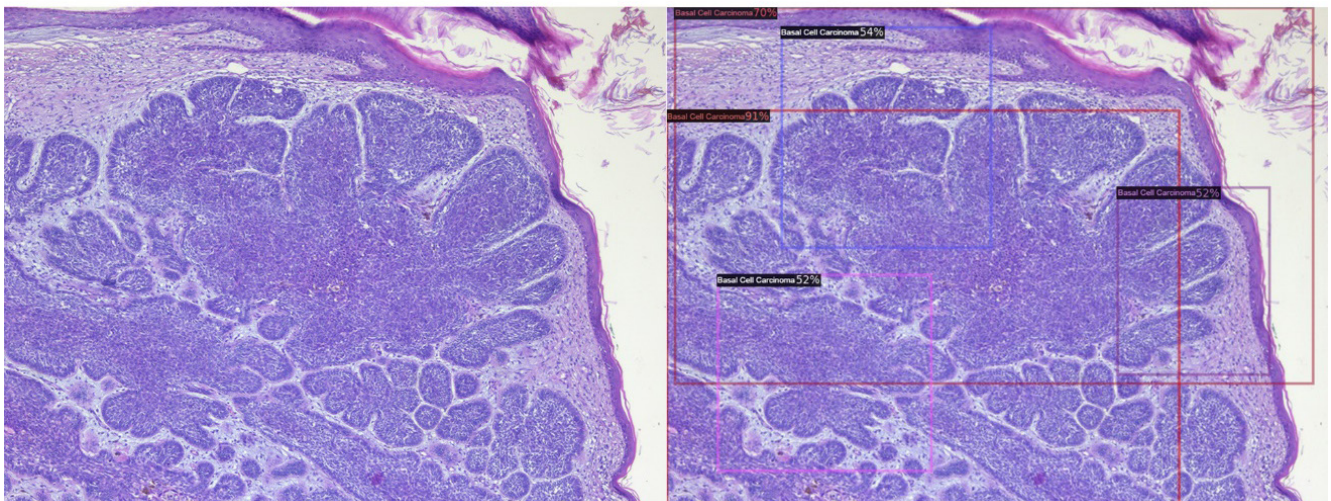

Fig. 3. Comparative presentation of a histopathological image, before processing (left side) and after processing (right side), Basal cell carcinoma (cBCC), HE, 20x

polygons detected by the algorithm and the marked poly- gons, calculating the sum of the masks, and dividing the
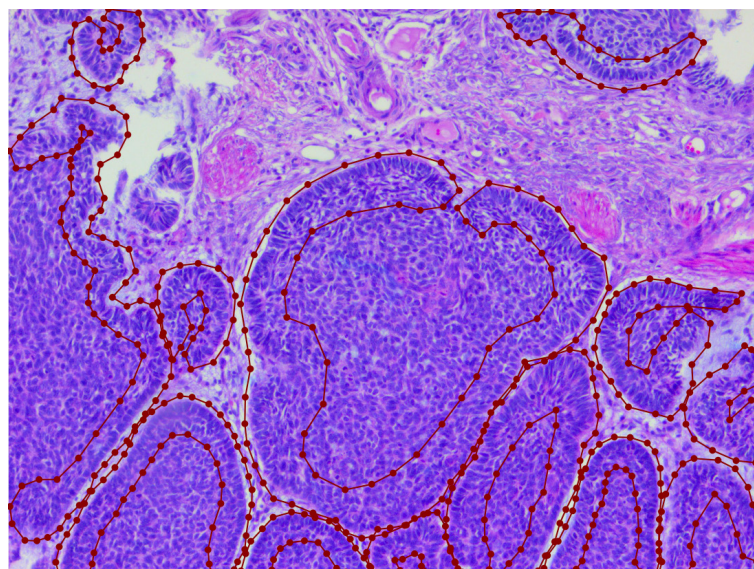

Fig. 4. Labeled histopathological image, red lines represent marked areas with pathological changes, Basal cell carcinoma (cBCC), HE, 20x

sum of the elements with value 2 by the elements with value greater than 0. The DSC was obtained similarly by multiplying the intersection of the values with value 2 and those with the value 0 and dividing the number obtained by the sum of the number associated with the number of pixels in both masks.

The Mask RCNN model successfully identified in 85% of cases the areas with pathological changes, the images that raised problems identifying pathologically changed cells are the slides with a color variation, turning towards a pale pink due to the discoloration of the slides over time or the difference given by the coloring technique used at the time of preparation of the respective slide. An example can be seen in Figure 5.

Figure 6 illustrates the differences between output of the program before training the algorithm with training dataset and after the learning process and the output of the algorithm after the training process. We can see that the obtained image from the algorithm before being trained is black and white and the algorithm was not able to detect a pattern representing pathological modifications specific to BCC. After the training we can clearly see the improvements. The image has the specific HE coloring and also the machine learning model was able to discover a pattern for the specific pathological modifications of BCC.

Practical integration
To demonstrate the utility of our study, we integrated the AI model into an Application Programming Interface (API). The API was intended to assist qualified medical personnel in the establishment of a definitive histopathological diagnosis of BCC. At the same time, this practical integration of the program helps the doctor by recording the pathological history of each patient. The way this web application works can be seen in Figure 7. The created algorithm and API were implemented using Flask (Python).

The server allows the user to create an account, log in, save, and view their own history and, last but not least, to upload data, which is processed by the program, and the user receives as a result, a .PNG file. All data is saved in a SQL database.

Figure 7 presents the main pages of the application, and we can see that when the cursor touches the icon intended for the user on the top right of the screen, the site displays the user's details and two buttons: "Sign out" button that will log the user out of the account alongside the "History" button.

Discussions
Once computational pathology is used by a laboratory, there are a multitude of new opportunities based on digital systems that increase workflow efficiency and facilitate diagnosis with specialized applications.

These applications, based on AI and deep learning, have proven to reduce error rates in establishing the diagnosis and increase the quality of the medical services provided [11].

Over the past decade, the introduction of AI into medical field has been increasingly evaluated, predominantly receiving favorable reviews, leading researchers to believe that it may benefit from wider use soon, particularly in medical imaging fields. There are hospitals in which specialties like radiology, have been equipped by manufacturing companies with devices that use AI and are designed
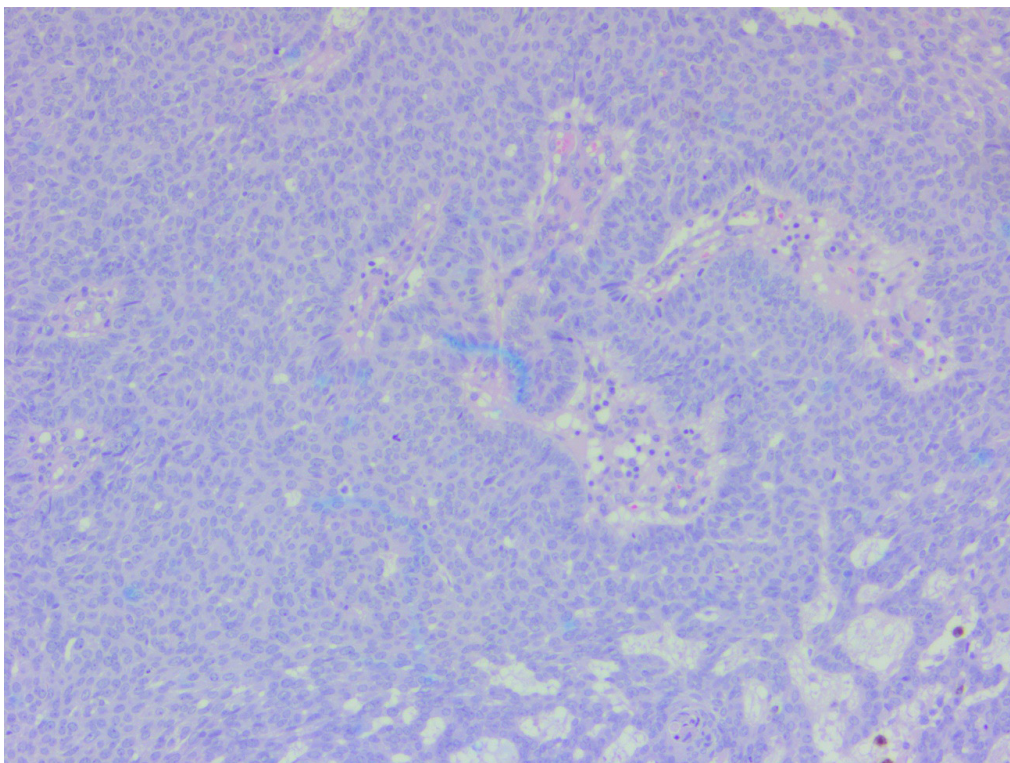

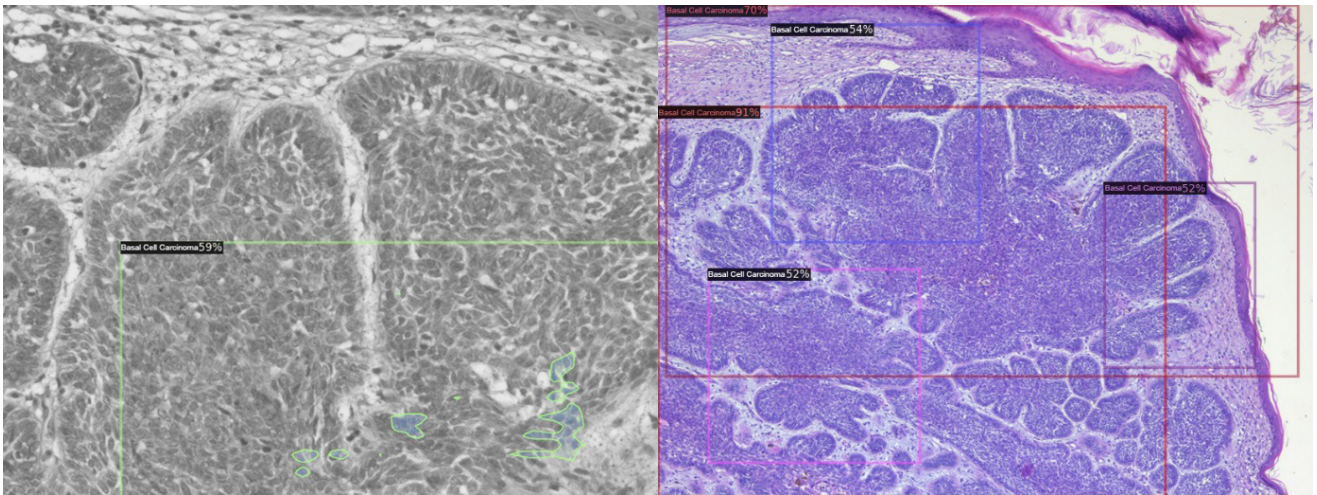Fig. 5. Example of cutaneous Basal cell carcinoma (cBCC), HE, 20x

Fig. 6. Before and after the training process, cutaneous Basal cell carcinoma (cBCC), 20x

Fig. 7. Web-based user interface of the application.

to answer specific routine diagnostic questions, and even though their effectiveness has been statistically demonstrated, none of these products are used nationally in any country in the European Union yet. In pathology, this transformation process has been further avoided, as in present whole-slide scanners are rarely used for routine diagnosis. Reasons range from high investment costs to uncertainty about data security, to reservations and reluctance among pathologists [12,13].

However, some medical institutions have decided to support digitization and the routine use of telepathology and even AI. In pathology institutes in Leeds, Utrecht, Pittsburgh and New York, digital interpretation has already been implemented in some or all cases and some companies already offer systems created specifically for Pathology departments.

In addition to scanning and digitizing whole slides and making a presumptive diagnosis, digital processing includes automatic barcode-based collection of case numbers, speech recognition-assisted dictation of results, and automatic transmission of findings to the hospital database [14].

The limitations in this field are: (1) the lack of pathologists with experience in the digital field, hence the apprehension in using these programs; (2) the increased complexity and weight regarding the management and integration of data from different sources to maximize patient care because in countries like Romania there is no centralized database of patients, data on their clinical and paraclinical status exists only fragmented; and (3) the learning process of these AI systems requires time and trained people, also the machine learning algorithms must be created in such a way that they are able to process and understand very large volumes of databases.

The method we chose to use in our study to perform the segmentation of histopathological images is Detectron2 developed by Facebook. We made this choice because Detectron2 is considered the gold standard in many Machine Learning situations, including instance segmentation. More precisely, the model chosen by us is represented by a Mask-RCNN model trained on Common Objects in Context (COCO) databases.

We trained the Detectron2 model using the data described in section 2.1 – Building the database, using 1000 epochs, with a learning rate of 0.00025 and a single classification category (ex: basal cell carcinoma, in this case). As a response, the algorithm provides the highlighting of areas that contain, with a certain probability, pathologically modified cells, specific for basal cell carcinoma. These images are then saved in .PNG format, making them easy to view on any device.

For fulfilling the purpose of this study, we used a number of 258 images obtained from cases diagnosed with basal cell carcinoma, cases extracted from the archive of the Pathology service within the Mureș County Clinical Hospital. Images were acquired using a Zeiss Axio microscope and ZenPro 3.2 image acquisition software. All obtained images have a resolution of 2560x1920 pixels.

Tumor cells were manually labelled by creating polygons both around them and around areas with a specific appearance of basal cell carcinoma, using the labelme library of the Python programming language.

For integrating the images into the program, we used the percentage rule of 80% of the images for the learning set (207 images) and 20% of the images for the test set (51 images).

To facilitate access to the AI model, we integrated it into an API (Application programming interface), which consists of 5 pages: the start page, the quick analysis page, the login page, the user page, the history of the analyzed images. The main advantage of integrating the model into a web application is the possibility of being accessed from any pathology department, anywhere in the world.

The design is minimalist, incorporating simple animations, for a pleasant user experience. The font and colors stand out, with the goal of encapsulating a pleasant, intuitive experience and simplifying the way this application is used by the public.

The application consists of 5 pages: Home page - represents the first page with which the user has contact when accessing the site; Quick Analysis Page; Login and registration page; Results page; History page.

Within the first page, there is a button to the analysis page, which offers the possibility of a quick analysis of an image. It should be noted, however, that if the user is not logged in, the analyzes performed are not saved, and the user will not be able to access them later. Near to the "Quick Analysis" button, we find the "Login" button, which can be used to redirect us to the login page. By pressing the button where we find a pencil illustrated, the user has the possibility to register, if he does not already have an account. To create an account, we must enter an email address and a password. After pressing the "Next" button, a pop-up window will appear and let us know if the process was completed successfully or not. After completing the registration process, the user has the possibility to log in. After successful login, the analysis page will be displayed.

From this point, the user has the option to upload an image of their choice. Now the application supports images in .jpg or .jpeg format. After uploading the image, the user has the option to delete the image, in case it was not chosen correctly. By pressing the "Analysis" button, the image will be sent for processing. During processing, a waiting page will be displayed. As soon as the result is delivered by the server, the user can view the analysis result. At this point, the result in .png format can be downloaded as an archive by pressing the "Download result" button.

This page displays a table containing the details of all the images analyzed by the user in his account: the date the image was processed, the name of the image and a link to the obtained result.

In recent years, AI has also begun to be used in the medical field to improve patient care by speeding up diagnosis and treatment processes. Most frequently, the concept of AI is used in medical specialties that work with the interpretation of images such as: radiology, pathology, but also electronically uploaded patient records can be evaluated by machine learning [15-17].

Since technological progress in the medical field aims to quickly establish a precise diagnosis and an individualized treatment, computational pathology is a major factor in the process of achieving this goal [18-20].

As more labs adopt digital pathology and digital slides become part of the normal workflow of diagnostic anatomic pathology, there will be better case management and case outcomes due to overall improved information management. The digital pathology workflow will help improve the efficiency of case interpretation, but as in other fields, the presence of the pathologist is essential in this AI-physician system [21-23].

Worldwide only a few pathology laboratories have switched to a fully digital workflow, and this could represent a blockade in the implementation of AI models in routine clinical diagnosis. Hybrid approaches, including the integration of Augmented Reality (AR)/Virtual Reality (VR) might be important intermediate steps [24].

The strategies to introduce AI in this medical field reported so far are based in the first place on the supervised learning method. This process involves that each of the images presented to the program during the learning process must first be annotated by a medical expert (pathologist). Since effective learning requires the input of thousands of example images, this leads to considerable effort and time consumption. The semi-supervised or unsupervised learning method could represent an answer to this problem.

Multi-instance learning is a technique where initially all the digitized slides are classified and afterwards only the regions with the lowest classification error are utilized. It is not necessary to annotate certain areas of the image; only the entire digitized slide is labeled, e.g. tumor yes/no. Using this technique and a total of 44,732 slides from 15,187 patients, Campanella et al. achieved AUROC values as high as 0.991 in the diagnosis of prostate cancer in a retrospective study. The previously discussed implications for establishing routine clinical diagnosis are of interest: if a system of this kind would be used for 65% to 75% of the pathology samples received, the sensitivity would be close to 100% [19,25].

Conclusions

The main areas of applicability of AI in health systems are those involving the manipulation of images to establish a diagnosis, because in this case deep learning and artificial neural network procedures can be more easily applied. The model used in the study, after a learning period, was able to successfully identify in 85% of cases the areas with pathological changes, the only problems encountered were

cell color variation. Using an already existing database, it is possible to generate models for rapid analysis of digitized slides, which gives the opportunity to solve cases in a shorter time with a lower error rate, and at the same time is useful for augmentation of the database and increasing the percentage of certainty The positive results of medical units where digitization is a routine process, support the introduction of AI in healthcare services as a mechanism to assist medical personnel.

Authors' contribution

ACD: conception and design, acquisition of data, analysis and interpretation of data, drafting the article, final approval of the version to be published

IGC: revising the article, acquisition of data, analysis of data, final approval of the version to be published

OSC: drafting the article, interpretation of data, final approval of the version to be published

LA: conception and design, acquisition of data, drafting the article, final approval of the version to be published

Conflict of interest
None to declare.

Bibliography

1. Amisha, Malik P, Pathania M, Rathaur VK. Overview of AI in medicine. J Family Med Prim Care, 2019;8(7):2328-2331.
2. Hamet P, Tremblay J. AI in medicine. Metabolism, 2017;69S:S36-S40.
3. Mayo RC, Leung J. AI and deep learning - Radiology's next frontier? Clin Imaging, 2018;49:87-88.
4. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using AI to Reduce the Risk of Nonadherence in Patients on Anticoagulation Therapy. Stroke, 2017;48(5):1416-1419.
5. Yu KH, Beam AL, Kohane IS. AI in healthcare. Nat Biomed Eng, 2018;2(10):719-731.
6. Krittanawong C. The rise of AI and the uncertain future for physicians. Eur J Intern Med, 2018;48:e13-e14.
7. Cocuz IG, Cocuz ME, Niculescu R, et al. The Impact of and Adaptations Due to the COVID-19 Pandemic on the Histopathological Diagnosis of Skin Pathologies, Including Non-Melanocyte and Melanoma Skin Cancers-A Single-Center Study in Romania. Medicina (Kaunas), 2021;57(6):533.
8. Krittanawong C. Healthcare in the 21st century. Eur J Intern Med, 2017;38:e17.
9. Bi WL, Hosny A, Schabath MB, et al. AI in cancer imaging: Clinical challenges and applications. CA Cancer J Clin, 2019;69(2):127-157.
10. Liang M, Tang W, Xu DM, Jirapatnakul AC, Reeves AP, Henschke CI, Yankelevitz D. Low-Dose CT Screening for Lung Cancer: Computer-aided Detection of Missed Lung Cancers. Radiology, 2016;281(1):279-88.
11. Peris K, Fargnoli MC, Garbe C, et al. European Dermatology Forum (EDF), the European Association of Dermato-Oncology (EADO) and the European Organization for Research and Treatment of Cancer (EORTC). Diagnosis and treatment of basal cell carcinoma: European consensus-based interdisciplinary guidelines. Eur J Cancer, 2019;118:10-34.
12. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of AI technologies in medicine. Nat Med, 2019;25(1):30-36.
13. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of AI in medicine. Wiley Interdiscip Rev Data Min Knowl Discov, 2019;9(4):e1312.
14. Försch S, Klauschen F, Hufnagl P, Roth W. AI in Pathology. Dtsch Arztebl Int, 2021;118(12):194-204.
15. Matheny ME, Whicher D, Thadaney Israni S. AI in Health Care: A Report From the National Academy of Medicine. JAMA, 2020;323(6):509-510.
16. Liu F, Zhou Z, Samsonov A, et al. Deep Learning Approach for Evaluating Knee MR Images: Achieving High Diagnostic Performance for Cartilage

Lesion Detection. Radiology, 2018;289(1):160-169.

17. Bartels R, Dudink J, Haitjema S, Oberski D, van't Veen A. A Perspective on a Quality Management System for AI/ML-Based Clinical Decision Support in Hospital Care. Front Digit Health, 2022;4:942588.

18. Parwani AV. Next generation diagnostic pathology: use of digital pathology and AI tools to augment a pathological diagnosis. Diagn Pathol, 2019;14(1):138.

19. Mandong BM. Diagnostic oncology: role of the pathologist in surgical oncology--a review article. Afr J Med Med Sci, 2009;38 Suppl 2:81-8.

20. Amin W, Srintrapun SJ, Parwani AV. Automated whole slide imaging. Expert Opin Med Diagn, 2008;2(10):1173-81.

21. Chen PC, Gadepalli K, MacDonald R, et al. An augmented reality microscope with real-time AI integration for cancer diagnosis. Nat Med, 2019;25(9):1453-1457.

22. Elmore JG, Longton GM, Carney PA, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA, 2015;313(11):1122-32.

23. Brimo F, Schultz L, Epstein JI. The value of mandatory second opinion pathology review of prostate needle biopsy interpretation before radical prostatectomy. J Urol, 184(1):126-30.

24. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med, 2019; 25(8):1301-1309.

25. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. AI in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol, 2019;16(11):703-715.