Deep Learning in Pulmonary Embolism detection

George Ionut Buleandra[1]

[1]Undergraduate Student at De Montfort, Leicester, UK

Artificial intelligence (AI) has garnered significant attention in recent years for its potential to revolutionize various fields, including medicine. In the field of radiology, AI has shown promise as a tool for aiding in the detection and diagnosis of diseases such as pulmonary embolism (PE). PE, a potentially life-threatening condition, is caused by a blockage in the pulmonary artery and is typically diagnosed using computed tomography pulmonary angiography (CTPA) scans. Early diagnosis is crucial, more than half of patients die within 2 hours of presentation. Given the high stakes and the increasing workload faced by radiologists, the development of AI algorithms for PE detection has attracted significant interest from researchers. In this context, convolutional neural networks (CNNs) have emerged as a promising approach, with some studies demonstrating their ability to perform on par with radiologists in detecting PE on CTPA scans. In this paper, we will review the current state of research on the use of AI for PE detection and discuss the potential benefits and challenges of incorporating AI into clinical practice.

Deep Learning | Pulmonary Embolism | Transfer Learning

Correspondence: ***buleandrageorge@gmail.com***

Introduction

Pulmonary embolism is a common and potentially deadly cardiovascular disease. It is the third most common form of this type of disease and has a high mortality rate if left untreated. According to Bělohlávek et al. (2013), more than half of patients with pulmonary embolism die within 2 hours of presentation, highlighting the importance of early diagnosis. Torbicki et al. (2008) as cited by Bělohlávek et al. (2013) affirm that CTPA (CT pulmonary angiogram) has become the gold standard for assessing patients for pulmonary embolism. However, the downside is that it takes a substantial amount of time to assess CTPA based on radiologists' expertise (Sofer et al. 2021). According to McDonald et al. (2015) as cited by Burs et al. (2021), it is estimated that a radiologist assesses an image every three seconds. Brady et al. (2017) as cited by Burs et al. (2021) also show that this massive workload can lead to a significant number of false negatives and misdiagnoses. As summarized by Burs et al. (2021), based on the work of Arbabshirani et al. (2017), Prevedello et al. (2017), and Chang et al. (2018), AI (Artificial Intelligence) has the potential to alleviate the burden on radiologists and improve the accuracy of diagnoses. Huhtanen et al. (2022) reached a similar conclusion, finding that an automated system for detecting PE could help radiologists avoid mistakes and prioritize the reading of positive cases for prompt evaluation.

Deep Learning in Pulmonary Embolism dectection

Artificial intelligence (AI) has been a promising area of research for the detection of pulmonary embolism (PE) in computed tomography pulmonary angiography (CTPA) images. According to Litjens et al. (2017), automated analysis of medical imaging has been an active area of research since the advent of computers, and by the end of the 1990s, such systems were widely used in medical imaging analysis. Sofer et al. (2021) conducted a survey of papers published since 2020 that evaluated the use of deep learning models for PE detection on CTPA images and found 7 papers met the inclusion criteria of being written in English, peer-reviewed, and containing measurable outcomes. These papers showed that convolutional neural networks (CNNs) may perform comparably to radiologists, with pooled sensitivity and specificity of 0.88 and 0.86, respectively, compared to sensitivities of 0.67 to 0.87 and specificities of 0.89 to 0.99 for radiologists. Tajbakhsh et al. were among the first to demonstrate the performance of CNNs in detecting PE in 2015, obtaining a sensitivity of 83% with two false positives per volume and the superiority of them over the classic machine learning techniques. Weikert et al. (2019) further improved the performance of CNNs through training on a large dataset of nearly 30,000 CTPAs, achieving a sensitivity and specificity of 92.7% and 95.5%, respectively. The performance of other models in the study is summarized in Table 1 (Sofer et al., 2021)."

The missing key between reliable product and proof of concept.

Convolutional neural networks (CNNs) are often trained and validated on the same dataset, which is typically split into training, validation, and testing samples. However, this type of validation does not meet the criteria for clinical validation, which requires data from multiple institutions to better represent the range of patient presentations and avoid bias and overfitting of the model. Overfitting may also occur due to small dataset sizes, as noted by Mazurowski et al. (2019), who pointed out that datasets used to train CNNs are often much smaller than the millions of samples typically used due to the cost and scarcity of expertly labeled data. This can lead to a lack of confidence in the model's generalization ability and a low level of confidence in its performance. Researchers have also emphasized the importance of validating the model, as the output of a CNN cannot be easily explained due to its "black-box" nature (Huhtanen et al. 2022, Sofer et al. 2021, Park et al. 2018).

Testing in real time workflow.

Recent research has demonstrated that CNNs can be seamlessly integrated into real-time workflows and perform on par with radiologists. Buls et al. (2021) tested the accuracy of Aidoc version 1.3, a FDA-approved CNN-based tool, in a real-world radiology setting at Universitair Ziekenhuis Brussel. The model had an accuracy of 98% and a substantial concordance with expert readings, as indicated by a kappa value of 0.78. The false positive rate was 24%, largely due to factors that made the readings challenging for both the AI and experts, such as body anomalies, comorbidities, and imaging artifacts. The authors concluded that the tool has the potential to act as a "second reader" for non-contrast CT and CTPA scans in detecting intracerebral hemorrhage and PE, respectively. Cheikh et al. (2021) reached similar conclusions in their study of a commercialized AI algorithm called AIDOC 1.0, which is based on a CNN and was tested in Imadis, a remote emergency radiology service in France. The researchers compared the performance of the AI and radiologists using three cohorts from 2018 to 2020 and found comparable results for both.

Beyond and above in detecting Pulmonary embolism.

The RSNA Pulmonary Embolism CT Dataset, comprising 9000 CTPAs carefully selected by the Radiological Society of North America and the Society of Thoracic Radiology from institutions in five countries (COLAK et al., 2021), presented a valuable opportunity for researchers to assess the performance of CNN models on a reliable database. Ma et al. (2022) used this dataset to demonstrate that CNNs can not only predict the presence of PE, but also the location (left, central, or right) and type (chronic or acute) of PE. The AUROC values for these predictions ranged from 0.95 to 0.69, as detailed in Table 2 (Ma et al. 2022). In the same study, the authors showed that the "black box" nature of CNNs can be partially overcome by using techniques such as gradient-weighted class activation mapping (Grad-CAM) and label-specific attention heatmaps, an example of which is shown in Figure 1 (Ma et al. 2022).

Prediction Parameter	AUROC
Negative Exam for PE	0.93
Intermediate	0.86
Chronic	0.69
Acute and Chronic	0.86
Central PE	0.95
Left side PE	0.89
Right side PE	0.92
R. ventricle L. ventricle radio greater than 1	0.87
R. ventricle L. ventricle radio less than 1	0.85

Table 2. AUROC values for predicted labels

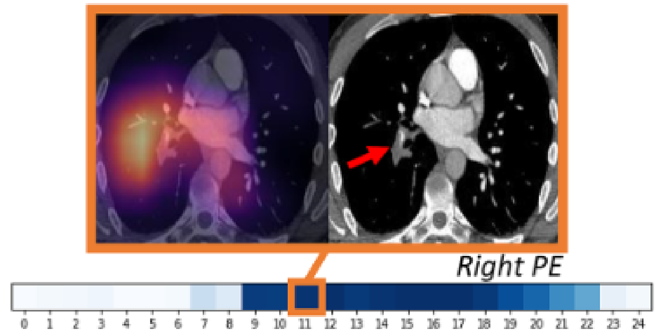

Fig. 1. Heatmap. Left side prediction of the model. Right side, annotated CTPA by radiologist

Transfer Learning.

Transfer learning plays a crucial role in training these models because it requires substantial amounts of data, which may be expensive and difficult to collect as presented above. The models are usually initially trained on a dataset of natural images, which is more accessible and can include millions of samples. Then they are trained further on the available CTPA dataset (Litjens et al. 2017). Menegola et al. (2016) as cited by Litjens et al. (2017) compares the performance of pre-trained networks and trained from scratch networks on 1000 samples and discovers that pre-trained performed better, but the study is too small to be significant.

Radiologists' Opinion

Radiologists have generally been receptive to the incorporation of artificial intelligence (AI) into their practice, finding it helpful in supporting their diagnoses and reducing workload. Cheikh et al. (2022) surveyed 79 radiologists and found that 55 "positively evaluated the AI algorithm to improve their diagnostic comfort." The authors concluded that "AI for PE detection appears to be a safety net in emergency radiology practice due to high sensitivity and NPV, thereby increasing the self-confidence of radiologists." A separate survey of 1032 radiologists in Italy in 2019 found that the majority believed that AI would decrease the error rate (750/1027) and optimize their work (697/1027, 67.9%). Most were not worried about AI replacing them (917/1032, 88.9%), and even more (77%) were in favor of adopting AI, with only 5% against and the rest uncertain. The only concern among the radiologists was that AI might damage their professional reputation (Coppola, 2021).

Conclusions

Based on the research discussed, it is clear that artificial intelligence (AI) has the potential to be a valuable tool in the detection and diagnosis of pulmonary embolism (PE). Several studies have demonstrated that convolutional neural networks (CNNs) can perform on par with radiologists in detecting PE on computed tomography pulmonary angiography (CTPA) scans, and some have even shown that AI can predict the location and type of PE with high accuracy. The RSNA Pulmonary Embolism CT Dataset has provided a reliable database for researchers to test the capabilities of CNNs in PE detection, and techniques such as gradient-weighted class activation mapping (Grad-CAM) and label-specific attention heatmaps can help to reduce the "black box" nature of these models. Overall, radiologists have shown openness to the use of AI in their work, with many finding it helpful in supporting their diagnoses and reducing workload. While some concerns have been raised about the potential for AI to diminish radiologists' professional reputation, the majority of radiologists surveyed were in favor of adopting AI and did not view it as a threat to their job security. However, it is incredibly important to understand the limitations of these models as they can critically impact patient care.

Functional Requirements

The software's purpose is to support radiologists in diagnosing Pulmonary Embolism by offering a starting point by highlighting some areas of the CTPA that may represent pulmonary embolism. To achieve this, the product must have a way of uploading the CTPA, displaying the CTPA, and highlighting the pulmonary embolism if it is the case, with a level of confidence and its position. Another important feature is to create a report with the results of the process. The product will have a single user type, the radiologist.

Use case 1: prediction of presence of pulmonary embolism based on CTPA.
Actor: Radiologist
Basic Flow: The radiologist loads the CTPA into the software. The software displays the CTPA with highlighted areas found as pulmonary embolism. Also, the number of the CTPA is displayed, presenting the prediction and the confidence level of the prediction. The "Generate Report" button becomes available. The radiologist generates the report for CTPA pressing the button.
Alternative flow A for case 1: The CTPA loaded does not meet the requirements
Flow: The radiologist loads the CTPA into the software. An error message is displayed describing the problem.

Table 3. Use cases. Contains all the use cases of the software

A further discussion is needed regarding this product as it can be mistakenly used as a Medical Device conform European Commision, detailed in the document "Guidance document Medical Devices - Scope, field of application, definition - Qualification and Classification of standalone software - MEDDEV 2.1/6" (Available at: https://ec.europa.eu/docsroom/documents/17921. Accessed 6/01/2023). The software product does not qualify as a Medical Device as defined in the guidelines because:

1. "... represents a set of instructions that processes input data and creates output data." The input data is represented by the CTPAs, the most important set of instructions is represented by the model that makes the prediction, and the output data is represented by the prediction and highlighted regions on CTPA.

2. " ... is not incorporated in a medical device at the time of its placing on the market or it's making available." It comes as standalone software able to run on any machine with a Windows OS.

3. "...is performing an action data different from storage, archival, communication or simple search." The action is represented by the calculation of the probability that the CTPA examined presents a case of pulmonary embolism.

4. "...the action is for the benefit of the patients". The product's purpose is to reduce diagnosis time for suspected patients of pulmonary embolism.

5. "...it is for the purposes defined in art 1.2(a) of Dir. 93/42 CEE". The purpose of the product is NOT the diagnosis of disease, but as define art 1.2(e) of Dir. 93/42 CEE" is "...intended for use by a duly qualified medical practitioner when conducting investigations as referred to in Section 2.1 of Annex X in an adequate human clinical environment." thus qualifies as a "device intended for clinical investigation".

Because it is not medical device and it can mistakenly being used as one, it must include a disclaimer which states the purpose of the product and how it can be safely used.

The following Use Case Diagram offers a better view of the functionalities of the software.

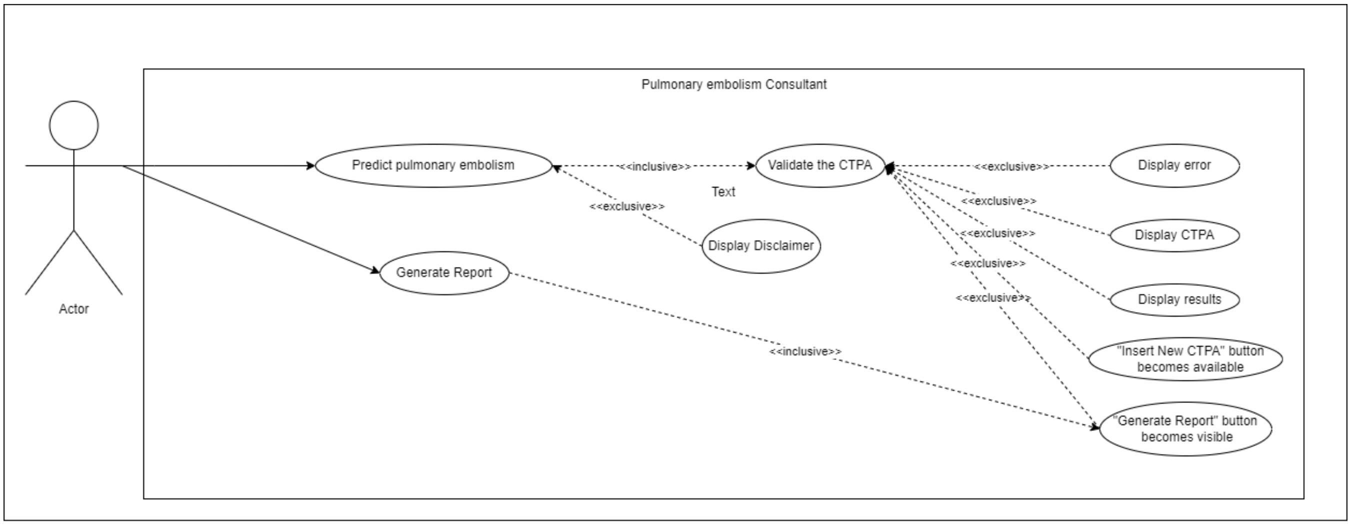

Fig. 2. Use Case Diagram Left side prediction of the model. Right side, annotated CTPA by radiologist

Functional Requirements are:

1. provide a way of uploading a CTPA in the software
2. display the current CTPA
3. highlight the areas which are identified as pulmonary embolism
4. provide a level of confidence of the prediction
5. provide a position of the embolism
6. generate a report with the annotations made on the CTPA and with the results of the processing.
7. display a disclaimer every time the application is started.

System design and documentation

The architecture of the computer applications is based on Model View Controller (MVC). This approach eases the application's management by splitting the functional requirements into 3 sections. The "View" refers to the part of the application that the user interacts with. The Model refers to database used, the format in which the data is stored. In this case the Model is represented by the Convolutional Neural Networks model. The controller represents the application's logic, is often called the "glue" between the view and the model. The purpose of it is to bridge the view to the model, in this case it is to validate the CTPA, then obtain the AI's prediction.

The system will be developed using Python 3 because it is a programming language which facilitates the fast development of applications through its concise syntax, solid set of libraries and community support. Python comes with a set of libraries for frond end such as: Pytq5, Tkinter, Kivy to name just a few. When it comes to back end, python is the choice of most programmers when prototyping and implementing AI (Artificial Intelligence) algorithms, having one of most popular set of the libraries: TensorFlow, PyTorch, Theano etc. All these features will create time for testing unique designs of product by reducing the time necessary for coding.

Front end

The front end must offer an intuitive design to facilitate the interaction with the model and the system's functionalities. The next elements for the front end are proposed to cover the functional requirements described above:

- CTPA frame with the annotations
- right side panel
  - the identification number of the CTPA
  - the probability of PE presence in the CTPA
  - the position of PE

- Generate Report Button

- Insert New CTPA button

A mockup of the interface can be seen in Figure 3.

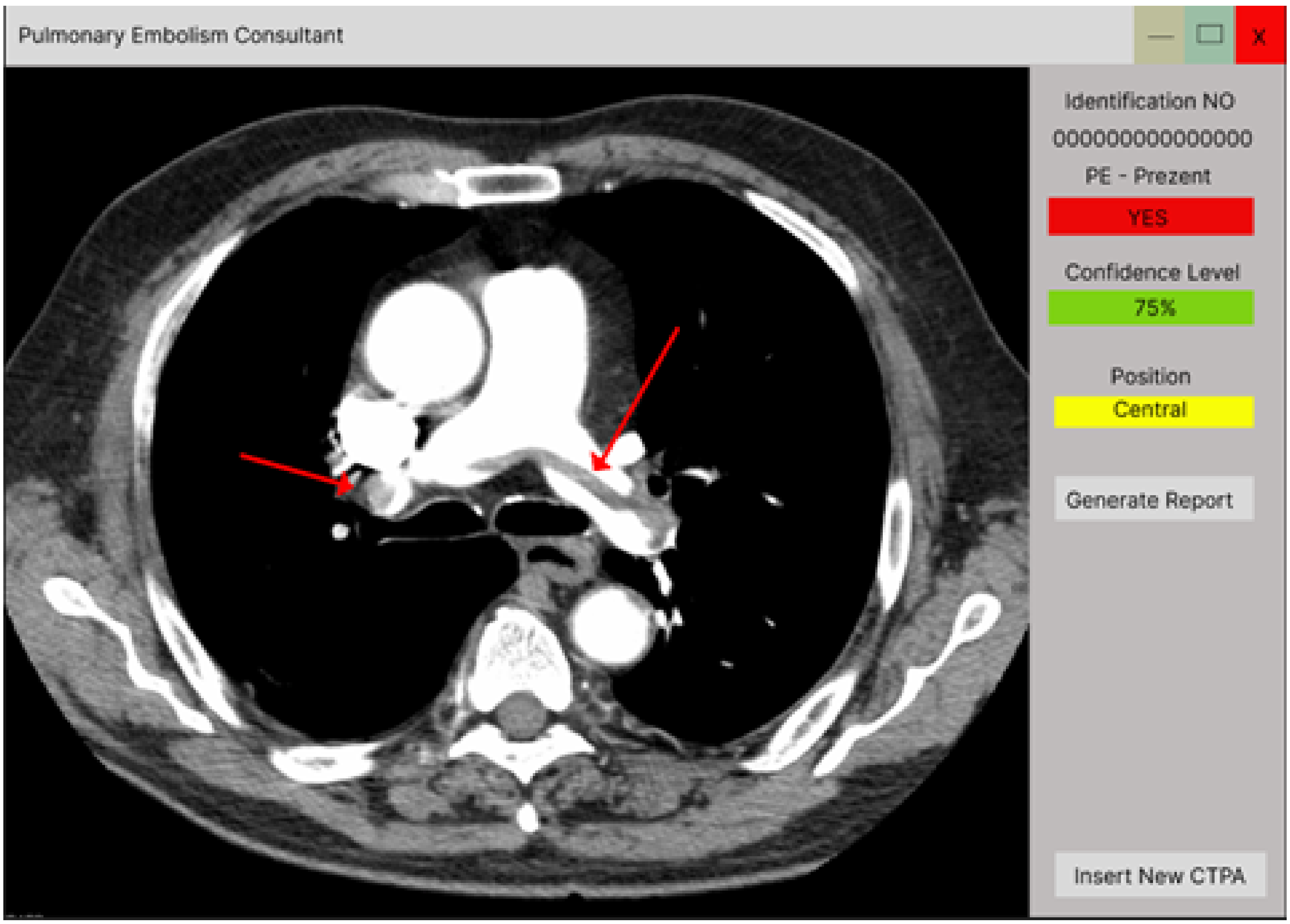

Fig. 3. Interface Mock-up

Back end

The back end is composed of controller and a convolutional neural network which predicts based on the CTPA the presence of pulmonary emboli. The controller retrieves the CTPA from the interface, validates the CTPA and if it is valid it is send for being processed by the AI model, else an error message will be send to the interface. The results of the processing of the AI model are retrieved by the model and send to the interface. Also, it will make the "Generate Report" available or unavailable accordingly. For a better understanding of the flow of the application the diagram in Figure 4 is proposed.

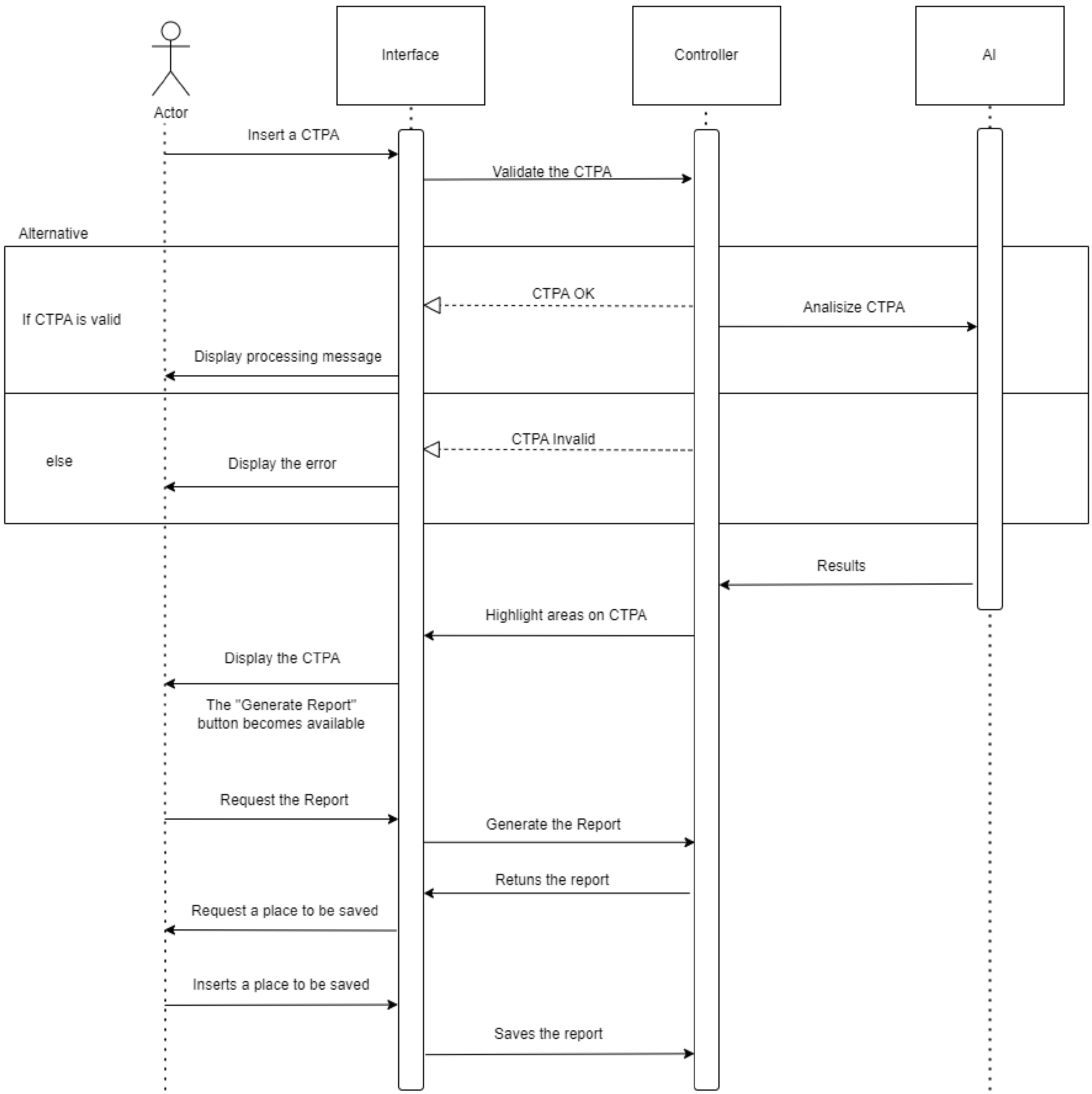

Fig. 4. Use Case Diagram Left side prediction of the model. Right side, annotated CTPA by radiologist

Indicative Test Plan

The testing of the software can be split into 2 parts. The first part covers the interface and the controller and the second part the AI model. In the first part the buttons of the interface will be tested, the displaying of the right messages. For the controller, the validation will be tested and how well the connection between the AI model and the interface is made. In the second part, the AI models will be trained and tested in parallel to discover which model and configuration offers the best performance.

Test no	Test Case objective	Test case description	Expected	Passed	Comments
1	The CTPA can be loaded in the software using drag and drop	The CTPA is dragged and dropped in the window	The processing begins	-	Not tested yet
2	The CTPA can be loaded through the "Upload CTPA" button	The Button is pressed, a file explorer window	The button opens the file explorer	-	Not tested yet.
3	The processing starts after loading the CTPA	Select the CTPA and press "OK"	The processing should start.	-	Not tested yet
4	An info message is displayed	Load a CTPA in the application	A status message is displayed	-	Not tested yet
5	An error is displayed if the CTPA does not fit in the requirements	Load a CTPA that does not meet the requirements	An error is displayed stating the issue	-	Not tested yet
6	The disclaimer shows up when the application is started	Run the application, and check whether the disclaimer appears	A message box shows up.	-	Not tested yet
7	The id of the CTPA is updated after it was loaded in the application	Load a CTPA, check whether the id of the CTPA is updated in the application	The id displayed and the id of the CTPA coincide	-	Not tested yet
8	The prediction is updated	Load a CTPA	The text box is updated accordingly to the prediction	-	Not tested yet
9	The position is displayed in the textbox	Load a CTPA	The text box is updated accordingly to the prediction	-	Not tested yet
10	The "Generate Report" button becomes visible only after a CTPA was processed.	Check whether the button is visible before a CTPA to be loaded	No button appears without a CTPA to be loaded	-	Not tested yet
11	"Generate Report" creates the report with the exact values provided by the model.	Compare whether the values written in the report are equal to those given by the Model	The values must be equal	-	Not tested yet

Table 4. List of test cases

Implementation Report

The development started with the interface of the software. The PytQ5 came in handy due to its designer which helped to create the code based on the design created using it. The CTPA area, the labels for CTPA id, presence of PE and position where added. Also, the "Load CTPA" and "Generated Reported" were included. The features left to implement for interface are: the disclaimer, and to add colors for button's backgrounds and labels. The next step is to set up the environment for training the AI models. In the end to create the controller which manages transitions and transformation on interface and makes use of the AI model. The interface created so far can be seen in the figure 5.

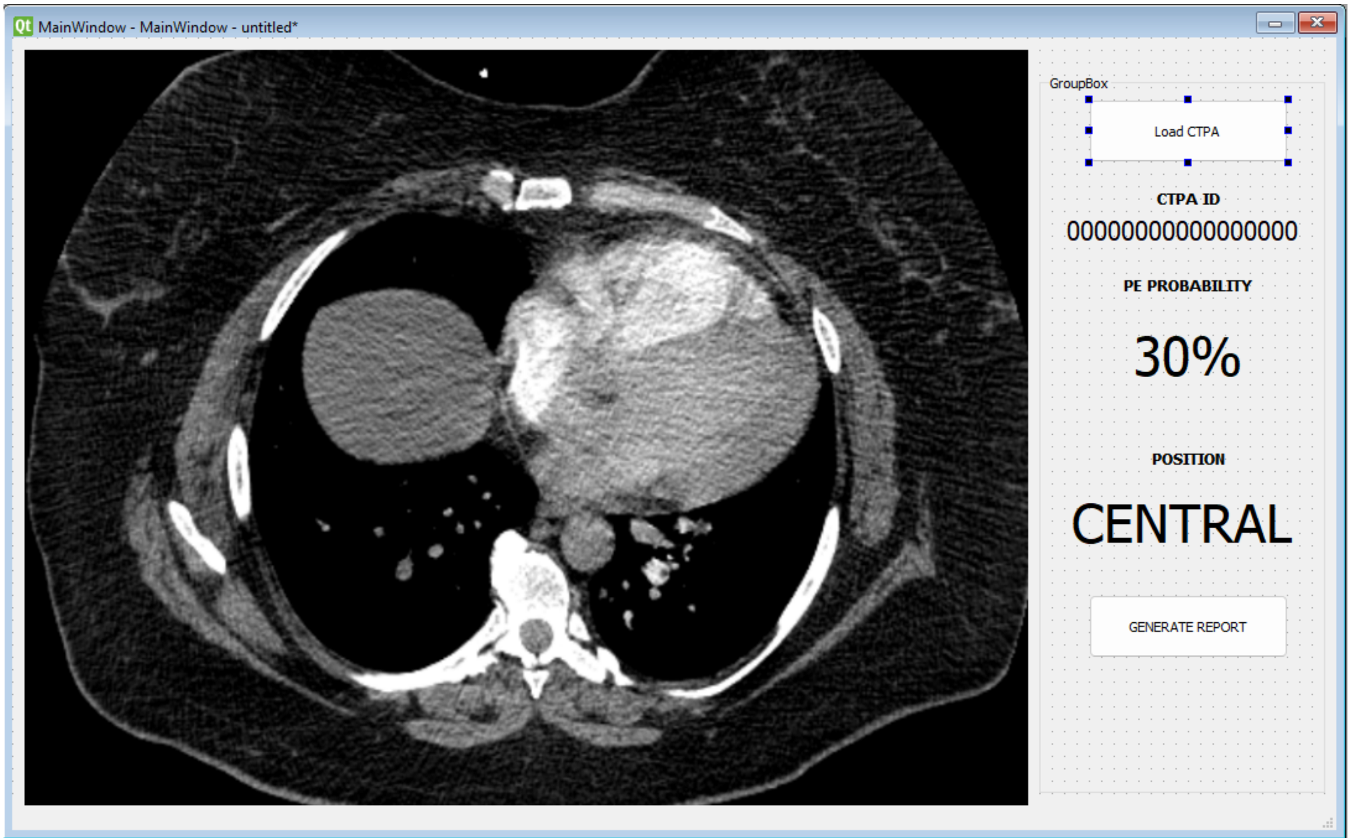

Fig. 5. Designed interface

ACKNOWLEDGEMENTS
I would like to thank to my supervisor, Dr Hossein Malekmohamadi, for his invaluable guidance that helped me in the vast domain of artificial intelligence.

Bibliography

BĚLOHLÁVEK, J. et al. (2013) Pulmonary embolism, part I: Epidemiology, risk factors and risk stratification, pathophysiology, clinical presentation, diagnosis and nonthrombotic pulmonary embolism.*Experimental and clinical cardiology*. [Online] 18(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3718593/ [Accessed 27/12/22].

BULS, N. et al. (2021) Performance of an artificial intelligence tool with real-time clinical workflow integration – Detection of intracranial hemorrhage and pulmonary embolism.*Physica Medica.* [Online] 83.Available from: 10.1016/j.ejmp.2021.03.015.[Accessed 27/12/22]

CHEIKH, A. B. et al. 2022, How artificial intelligence improves radiological interpretation in suspected pulmonary embolism.*European Radiology.* [Online] 32. Available from: https://doi.org/10.1007/s00330-022-08645-2.[Accessed 28/12/2022]

COLAK,E. et al. (2021) The RSNA Pulmonary Embolism CT Dataset.*Radiology: Artificial Intelligence.* [Online] 3(2). Available from: 10.1148/ryai.2021200254. [Accessed 28/12/2022]

COPPOLA, F. et al. (2021) Artificial intelligence: radiologists' expectations and opinions gleaned from a nationwide online survey.*Experimental and clinical cardiology* La radiologia medica [Online] 126. Available from: https://doi.org/10.1007/s11547-020-01205-y. [Accessed 28/12/2022]

HUHTANEN, H. et al. (2022) Automated detection of pulmonary embolism from CT-angiograms using deep learning.*BMC Medical Imaging.* [Online] 22(1). Available from: https://doi.org/10.1186/s12880-022-00763-z. [Accessed 27/12/2022]

LITJENS G. et al. (2017) A survey on deep learning in medical image analysis.*Medical Image Analysis.* [Online] 42. Available from: https://doi.org/10.1016/j.media.2017.07.005 [Accessed 27/12/22].

MA, X. et al. (2022) A multitask deep learning approach for pulmonary embolism detection and identification.*Sci Rep*[Online] 12. Available from: https://doi.org/10.1038/s41598-022-16976-9. [Accessed 27/12/2022]

MAZUROWSKI, M.A., et al. (2019) Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI.*J. Magn. Reson. Imaging* [Online] 49. Available from https://doi.org/10.1002/jmri.26534 [Accessed 28/12/2022]

Park SH, Kressel HY. (2018). Connecting Technological Innovation in Artificial Intelligence to Real-world Medical Practice through Rigorous Clinical Validation: What Peer-reviewed Medical Journals Could Do.*J Korean Med Sci.*[Online] 33(22). Available from: https://doi.org/10.3346/jkms.2018.33.e152 [Accessed 28/12/2022]

SOFER S. et al. (2021) Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis.*Sci Rep.*[Online] 11. Available from: https://doi.org/10.1038/s41598-021-95249-3 [Accessed 27/12/22].

Apendix

Author	Year	Datasize	Performance scores
Huang et al.	2020	1997	AUROC of 0.85 Sensitivity and specificity of 75% and 81%
Liu et al.	2020	878	AUC of 0.93 Sensitivity and specificity of 94.6% and 76.5%
Huang et al.	2020	1837	AUROC of 0.95 Sensitivity and specificity of 87.3% and 90.2%
Weikert et al.	2019	29,465	Sensitivity and specificity of 92.7% and 95.5%
Yang et al.	2019	129	Sensitivity of 75.4% at two false positives per volume
Rajan et al. (IBM)	2019	2420	AUC of 0.94
Tajbakhsh et al.	2019	121	Sensitivity of 83% at two false positives per volume

Table 1. The summery of models included in Sofer et al. (2021)'s survey on deep learning in pulmonary embolism detection, extras.