Bilkent University

Department of Computer Engineering

Senior Design Project

XAI-Empowered Breast Cancer Analysis

2023-2024 Fall Semester

Project Specification Document

Project ID: T2312

Project Team Member Names:

Ahmet Arda Ceylan	(22003148)
Mahmut Furkan Gön	(21902582)
Mustafa Yetgin	(21903191)
Omar Hamdache	(22001398)
Ömer Kağan Danacı	(21902584)

Advisor: Selim Aksoy

Contents

1. Introduction

1.1. Description

We have noticed the problems within the current healthcare system that result in hardships for individuals and institutions involved. The high cost of healthcare services in terms of both time and effort for patients, coupled with considerable error rates, make the existing healthcare system less effective. The growing population and diversity of diseases further complicate the ability of the current healthcare system to meet the demand. Additionally, inadequate working conditions for doctors have led us to develop a project to make improvements in this field.

Advancements in technology, such as robotic applications and machine learning techniques, have started to automate the healthcare system inevitably due to the advantages they offer. We are focusing on this transition process and aim to propose a transition model developed with novel technologies.

In our project, we have chosen to focus on breast cancer. Breast cancer accounts for 12.5% of all new annual cancer cases worldwide, making it the most common cancer in the world [1]. Especially it is more common among women. One out of every eight women has the possibility of having breast cancer [2]. Breast cancer can become fatal in late stages, comprising 15.5% of cancer-related deaths [2]. However, if it is detected in early stages and treatment begins, about 70-80% recovery rate has been observed [3]. Therefore, periodic controls are necessary to enable early diagnosis, particularly mammography is recommended once a year for women between the ages of 40 and 69 for breast cancer screening [2].

We propose an end-to-end system that is a doctor and patient assistant for early breast cancer screening. Empowered by XAI (explainable AI), it helps doctors, especially radiologists, to diagnose breast cancer at early stages with higher accuracy. The flexibility and accessibility that our system provides make it easier for patients to follow their screening results and contact their doctors. The availability of many doctors on one platform allows patients to access many opinions, satisfying the desire of the second opinion market.

In our system's workflow, patients upload their mammography result, which is analyzed by a doctor and our XAI model separately. Then, if the results match, the diagnosis will be finalized; otherwise, another doctor will review both results by taking advantage of the explainability of the model and give a terminal decision.

Our project has some key objectives. Firstly, we aim to make healthcare services more accessible and save patients valuable time and effort. We also have a goal of minimizing the error rate in the diagnosis process by taking advantage of the doctor and XAI collaboration. Finally, we aim to provide a much more flexible, comfortable, and profitable working environment for doctors.

In conclusion, our project is a pioneering work to revolutionize breast cancer screening and the broader healthcare system. By combining the expertise of medical professionals with cutting-edge technology, we aim to make a significant impact on early breast cancer diagnosis, benefiting both patients and healthcare providers.

1.2. High-Level System Architecture & Components of Proposed Solution

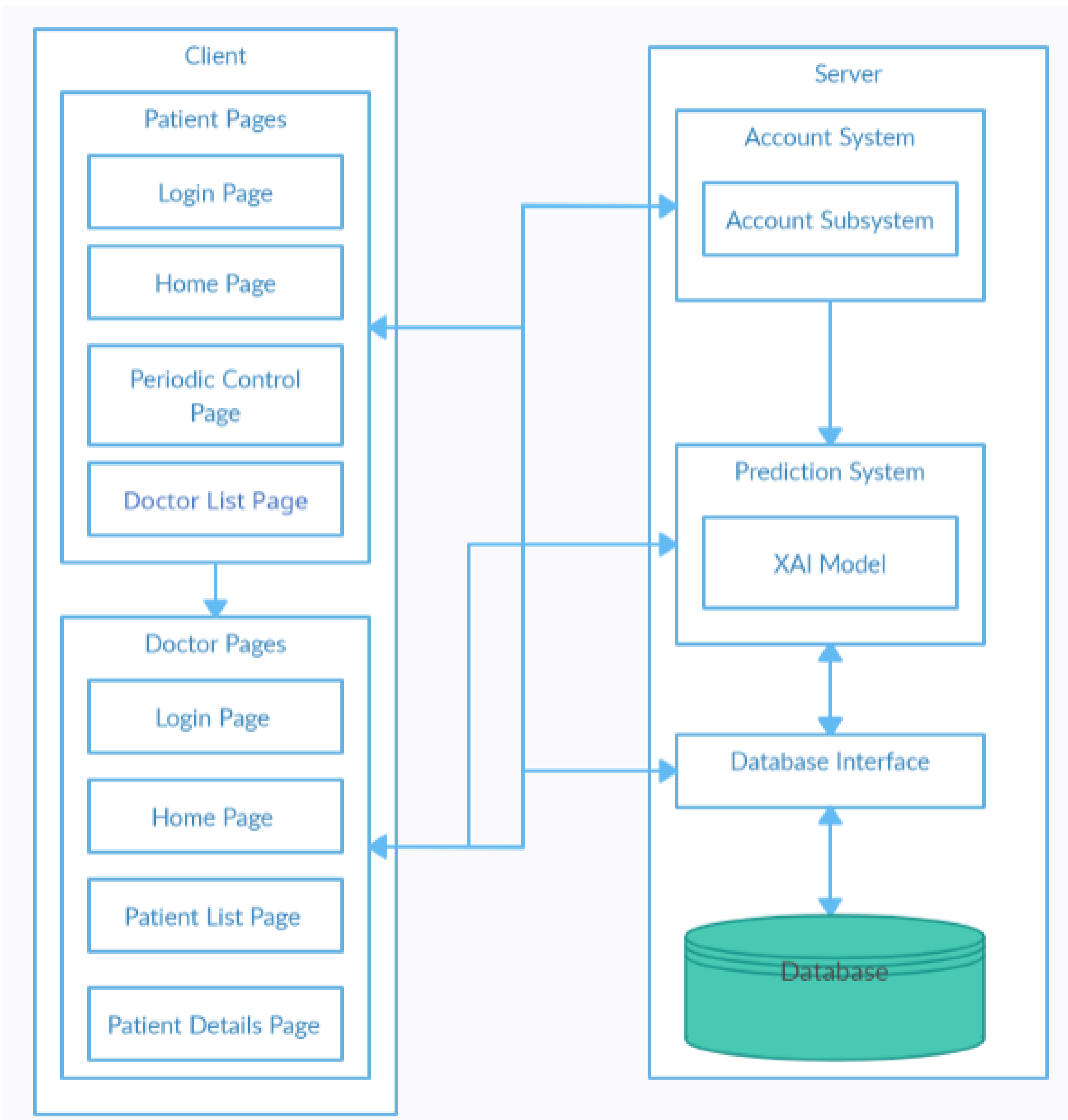

Fig.1: High-Level System Architecture

1.3. Constraints

1.3.1. Implementation Constraints

- The mammography data collected should be of high quality and consistency. Any noise or inaccuracies in the data can impact the model's performance and the doctor's evaluations. Furthermore, data that does not represent the variety in society has the risk of being biased and thus shows lower accuracy for those from different races and backgrounds than the rest of the society.

- To train the model with such big data, we should satisfy minimum requirements for both hardware and software tools.

- We should implement strong data security measures to protect patient information and comply with relevant data protection regulations like HIPAA.

1.3.2. Economic Constraints

- We should define a clear budget for the project, considering the costs of acquiring mammography data from data centers and compensation for the limited number of freelancer doctors at the beginning.

- With a clear budget at hand, we should balance cost and benefit by conducting a cost-benefit analysis to assess the economic feasibility of the project. We may consider the potential cost savings from early cancer detection and diagnosis.

- We should determine the possible sources of funding.

1.3.3. Ethical Constraints

- Patient's personal health information must be kept with the utmost care and compliant with relevant laws like the aforementioned HIPAA.

- Patients should provide informed consent for using their data in the project. They should understand the purpose and potential risks.

- We should be transparent about using machine learning in the diagnosis process and explain to patients and doctors how the model aids diagnosis. In addition, we should explain weak points in our model to doctors and tell them that the model is just an assistant in diagnosis with its explainable AI feature.

- The machine learning model should be used to support doctors rather than as a replacement for their expertise.

1.3.4. Legal Constraints

- We should consider who will take the legal responsibility for the misdiagnosis.
- The legal side of recruiting freelance radiologists should be regarded.
- Whether or not uploading user data to our system other than the legally predetermined system already used in KETEM and collaborative establishments is legally problematic will be investigated.

1.4. Professional and Ethical Issues

All the data that is accessed during the training of the machine learning model is obtained from publicly available databases that consist of medical images taken from patients who consented for their data to be used for medical and academic purposes. Additionally, some of the data is obtained by legal agreements with organizations' ethics commissions. Moreover, we will implement the model in a way that prevents the reconstruction of the training data or any part of it.

At the registration stage, consent of the users will be taken to process and store user data in our system. We will only keep the necessary data in our servers, and all other data will be stored in the user device as much as possible. All the stored data will be protected by encrypting them and securing our servers. The users will have the option to choose how long the data stays on our servers. Users are also asked for permission to use their data to further train our model (for feedback learning). We will authenticate the users before uploading any data to the system by checking their identity. We might cooperate with some companies that are specialized in the relevant field; for this, we will need to sign a Non-Disclosure Agreement (NDA).

Using machine learning models in critical domains like healthcare is a sensitive issue. The problem with classical models is that they act as a "black box"; it is not possible to understand the reasoning behind the classification. This poses an ethical issue since relying on a model that might misdiagnose a case without understanding how it made its decision might put patients' lives in danger. Therefore, we will include an XAI model to make it clear to doctors how the model diagnosed the case. This XAI model can provide text explanations and highlight relevant locations in the image.

2. Design Requirements

Design requirements are analyzed in two sections.

2.1. Functional Requirements

The following functional requirements are for the **patients**:

- The Turkish users must register/log in using their TR Identity No. The foreign user must register/log in using the Passport No.
- The user can upload/remove medical images in the .dcm format they want to be assessed by specifying when the image is taken.
- The user can see his/her already uploaded medical images.
- The user can request an assessment for diagnosis based on the selected medical images.
- The user can see the final diagnosis and take a report for that based on the images they uploaded.
- The user can see the optimal steps for the final diagnosis, such as making an appointment with an oncologist or requiring no further action.
- The user can see their old diagnosis decisions.
- The user can make payments for the medical image assessments with their preferred payment type and card.
- The user can add/remove the preferred payment type and card.

The following functional requirements are for the **doctors**:

- The user can register/log in using their username and password. (More would be required for registration, such as certificates.)
- The user can see the waiting medical image assessments.
- The user can access old images and diagnoses of the patient they assess.
- The user can see the explainable AI (XAI) diagnosis of the medical image they assess.
- The user can make an assessment and decide on the final diagnosis.
- The user can add notes and his/her suggestions to the final diagnosis and further steps.
- The user can see his/her accuracy rate compared to the XAI diagnoses.
- The user can see his/her income from assessing medical images.
- The user can add/remove bank account information.

The following functional requirements are for both **doctors** and **patients**:

- The user can edit his/her profile.
- The user can delete his/her profile.

2.2. Non-Functional Requirements

The non-functional requirements are analyzed under five sections.

2.2.1. Usability

The mobile application and web-based platform will feature an intuitive and user-friendly interface with a learning curve not exceeding 30 minutes for new users. Being a user-friendly application means users can achieve their planned operations as quickly as possible. For this reason, medical image uploading can be done in three clicks, and XAI explanations to the doctors will be accessible in two clicks. The system will support only the Turkish initially; however, English and Arabic languages are planned to be integrated into the system.

2.2.2. Reliability

The system is aimed to maintain a minimum uptime of 99% to ensure 24/7 availability. We will implement our system in a way that it is fault-tolerant. The system will implement robust encryption and access control mechanisms to ensure data integrity and prevent unauthorized access. We will integrate KPSPublic Web Service[4] into our system to enable users to register and sign in to our system with TR ID No and hence enhance security. The system will recover from failures within 5 minutes and maintain core functionality even with faults. The system will follow GDPR[5] and KVKK [6] regulations in private data management.

2.2.3. Performance

Medical image uploads, processing, and assessment will have response times of 2 seconds, 5 seconds, and 3 seconds, respectively. The system will support a 200% increase in concurrent users without a decrease in response time. The system will ensure a minimum data transfer speed of 10 MBps for medical image uploads. System resource utilization will be at most 70% of available CPU and memory to ensure optimal performance.

2.2.4. Supportability

This application will initially be developed as a cross-platform mobile application on the patient side and a web application on the doctor side. For the upcoming releases, the application scope can be extended as a web application on the patient side, and new features can be added. System updates and maintenance will be scheduled during off-peak hours to minimize disruption, and the process will not exceed 1 hour. The system will provide a full suite of documentation, including FAQs, troubleshooting guides, and system architecture documentation, accessible from the user interface for patients and doctors.

2.2.5. Scalability

The system will automatically scale resources to accommodate a 100% increase in concurrent users within 5 minutes. Load balancing mechanisms will evenly distribute incoming requests across available servers to prevent overload and optimize performance. We aim to develop our application so that our ML model is usable across many platforms, including other applications, such as integration to e-Nabız [7].

3. Feasibility Discussions

3.1. Market & Competitive Analysis

In Turkey, the breast cancer market size is $146 million in 2022 and is projected to reach $398 million in 2030 [8]. Each woman who is between 40 and 69 years old, which comprises about 14.3 million people, approximately 33.5% of the women in Turkey, according to the data from 2022 [9], needs to take periodic control once a year against breast cancer risk.

In Turkey, there is an institution called KETEM that provides breast cancer screening services. This governmental institution offers free mammography services for women who want to undergo breast cancer screening. However, the only role of this institution is to capture mammography images and transfer them to the facility where the medical analysis of the data will occur. Therefore, this institution is not a competitor for us. Nevertheless, it reflects the demand in this market.

The application has no direct competitors but indirect competitors. Those indirect competitors can be classified into two main application areas, which are the second medical opinion market and the doctor assistance market.

Online second medical opinion services enable patients to receive medical advice and evaluations from professionals through digital platforms. Patients upload their medical data, for instance, test results, for remote review by doctors in various medical fields. The doctor provides a detailed second opinion, including diagnosis confirmation and treatment recommendations, with communication occurring through secure messaging or video calls. These services are particularly beneficial for individuals seeking expertise not locally available and access to well-known professionals while potentially saving time and cost.

We have found two examples of online second medical opinion applications serving in Turkey, which are as follows:

- **Hastalar Soruyor:** Hastalar Soruyor is a web-based application that provides patients with several services, which can be listed as consulting a doctor, analyzing medical test results, getting a second doctor's opinion, and making an appointment with a doctor. For these services, patients are asked to make payment, which is determined before each service [10].

- **Medicopin.com:** Medicopin is a healthcare service that connects users with top medical experts worldwide, allowing them to share health concerns and receive expert evaluations via personalized video recordings or face-to-face consultations with doctors through Medicopin.com. It aims to provide users not only with reliable medical advice but also the peace of mind needed during severe health issues while offering ongoing support for a stress-free healthcare experience [11].

Doctor assistance products provide services for doctors to make the decision-making processes more straightforward. These products provide additional data about the sample by annotation, textual explanation, and framing of the possible abnormal regions in the medical data. Additionally, if there is an ML model integrated into the product the result of it is shared with the doctors.

We have found three different products that provide doctor assistance in mammography analysis and operate in Turkey:

- **Screen Point Transpara:** Screen Point Transpara is a product that aims to make early detection using machine learning methods for breast cancer. Additionally, it is FDA-cleared for use on all major 2D and 3D mammography systems. According to their research, they have a similar recall, cancer detection rate, and false positive rate compared to the standard double-checking. They currently provide services to the local private oncology clinics in İstanbul [12].

- **ICTerra:** ICTerra's effort in AI is framed in oncology and radiology. They have a project to classify mammograms using XAI. Their main target in the market is the governmental agencies and hospitals. The details of the project and benchmarks of the model are not shared publicly [13].

- **Siemens Syngo Breast Care:** Siemens is one of the leading companies in the imaging machine manufacturing industry. Additionally, they developed software that is integrated into the imaging mammography machine [14].

3.2. Academic Analysis

The technical feasibility of this project was analyzed from many aspects. The main areas of research were the availability of data, the accuracy of the ML model, the accuracy of an AI-Doctor hybrid diagnosis system, and finally, the applicability and accuracy of an XAI diagnosis system.

Firstly, it is necessary to have a large and diverse dataset for this project. Based on our research, the available data for breast cancer images (mammograms) is extensive and well-diverse. Therefore, we will make our data as unbiased as possible while maintaining its quality. The candidate datasets we are considering are The Cancer Imaging Archive datasets [15], which contain tens of thousands of images from multiple datasets; the King Abdulaziz University Breast Cancer Mammogram Dataset (KAU-BCMD), which has more than 1400 images [16]; and the Imaging Data Commons (IDC) National Cancer Institute containing more than 14000 cases [17]. The mentioned datasets will be enough to train the primary model and obtain good results from a generalized BI-RADS classifier. As a further step, it will be necessary to fine-tune the model to optimize it for early cancer cases since our goal is to implement a system that detects the cases from the screening stage. For this purpose, we will use a dataset from OMI-DB, which contains thousands of images from the UK [18]. These images are labeled by early and late cancer cases as well, which makes it possible to fine-tune the model for our purpose. A dataset from this database was used in the process of developing a machine-learning mammogram classifier for early breast cancer cases, and it even allowed for early breast cancer prediction. The results are depicted in the paper "International evaluation of an AI system for breast cancer screening" [19]. We might extend our dataset collection further based on the performance we observe from the model while training it. The current datasets are enough to train a fully working and accurate model, but in case of issues regarding the diversity of the data, we might include more datasets from different countries.

Then, we must ensure we can train a model that can give accurate results. The model must add value to the currently available system of only doctors. Therefore, the system must have high accuracy and reliability. Showing the feasibility of such a system is enough at this stage to proceed with the implementation. Many papers depict the accuracy of deep learning models, usually using a different variation of deep learning techniques. The primary deep learning technique that we will be using in our project is Convolutional Neural Networks (CNN). Lots of other variations of CNN show promising accuracy. For example, a comprehensive comparison between different CNN variations can be found in Table 2 of the paper under the title "Classification of Mammogram Images Using Multiscale all Convolutional Neural Network (MA-CNN)" [20].

Model	**Category**	**Specificity**	**Sensitivity**	***f*-score**	**AUC**
DCNN	Benign	0.80	0.84	0.82	0.93
	Malignant	0.86	0.88	0.87	0.97
	Normal	0.78	0.71	0.74	0.89
	Average	0.81	0.81	0.81	0.93
All-CNN	Benign	0.88	0.96	0.92	0.98
	Malignant	0.86	0.99	0.92	0.99
	Normal	0.97	0.74	0.84	0.97
	Average	0.90	0.90	0.89	0.98
Multiscale CNN	Benign	0.92	0.94	0.93	0.99
	Malignant	0.86	0.99	0.92	0.99
	Normal	0.95	0.79	0.86	0.97
	Average	0.91	0.91	0.90	0.98
MA-CNN	Benign	0.97	0.97	0.97	1.00
	Malignant	0.96	0.99	0.98	1.00
	Normal	0.96	0.94	0.95	0.99
	Average	0.96	0.96	0.97	0.99

Fig. 2: Table 2 of the MA-CNN paper showing a comparison in performance for different classes

Table 2 shows that MA-CNN is the most suitable variation of CNN since it has the higher AUC [Area Under the ROC (Receiver Operating Characteristic) Curve] as well as the best results regarding all other metrics [20, Table 2]. MA-CNN is a modified version of DCNN (Deep Convolutional Neural Networks) that was developed for mammography analysis. The idea is to modify the convolution and pooling layers, and the model is made to extract both low-level and high-level contextual features. MA-CNN combines both All-CNN and Multiscale CNN approaches, making it a more accurate approach for this task. The architecture that can be used for mammograms can be seen in Figure 3 of the same paper [20, Fig. 3]. The model comparisons can be seen in Table 1 [20, Table 1].

Model	Accuracy (%)	Error	# trainable parameters	Time per step (ms)
DCNN	81.13%	0.48	3,36,227	4
All-CNN	89.80%	0.30	3,14,851	6
Multiscale CNN	90.70%	0.28	5,83,971	22
MA-CNN	96.47%	0.12	4,82,787	20

Fig. 3: Table 1 of the MA-CNN paper showing general performance comparison [21]

Additionally, some other CNN techniques can be used to improve accuracy, efficiency, and other metrics. As an example, R-CNN and all the different inspirations from it can be explored for image analysis purposes [21].

The model that we are proposing to automate the periodic breast cancer check using mammograms has the combination of the two parties: doctors & ML model, as elaborated in Section 2. It is crucial to design the interaction between the two parties to make this model possible to be implemented in real life. Regarding this issue, as a group, we focused on the potential problems that could arise during this collaboration. According to research, it is found that doctors who get the advice (ML diagnosis in our case) are highly affected by the guidance during their decision-making process [22]. Additionally, it is found that having two different doctor analyses has a higher recall compared to a single doctor analysis [23]. Moreover, we decided to use an explainable machine learning model to decrease the possible bias that the doctors might have and make the collaboration between the two parties more transparent and smooth [24]. As a result of this literature review and the feedback from different doctors, we concluded to have this final model: first, both the ML model and the first doctor analyze the mammogram independently from each other; then, in case of a mismatch in their diagnoses, a second doctor checks the result and makes the final decision.

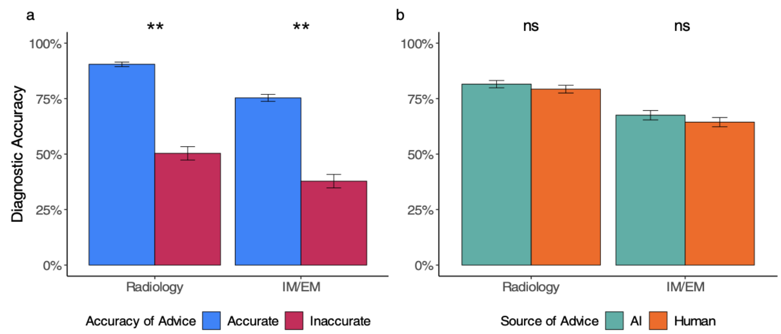

Fig. 3 Diagnostic accuracy across advice accuracy and source. We demonstrate the effect of the accuracy of advice and source of advice on diagnostic accuracy for task experts (radiologists) and non-experts (IM/EM physicians). In (**a**) we compare diagnostic accuracy across advice accuracy, demonstrating that both groups perform better when they receive accurate advice. In (**b**) we compare diagnostic accuracy across advice sources, demonstrating that neither group of physicians had a significant difference in diagnostic accuracy depending on the source of advice. There is no significant interaction between advice accuracy and advice source. The error bars represent confidence intervals. $^{*}p \leq 0.05$, $^{**}p \leq 0.001$, ns = not significant.

Fig. 4: Diagnostic accuracy across advice.

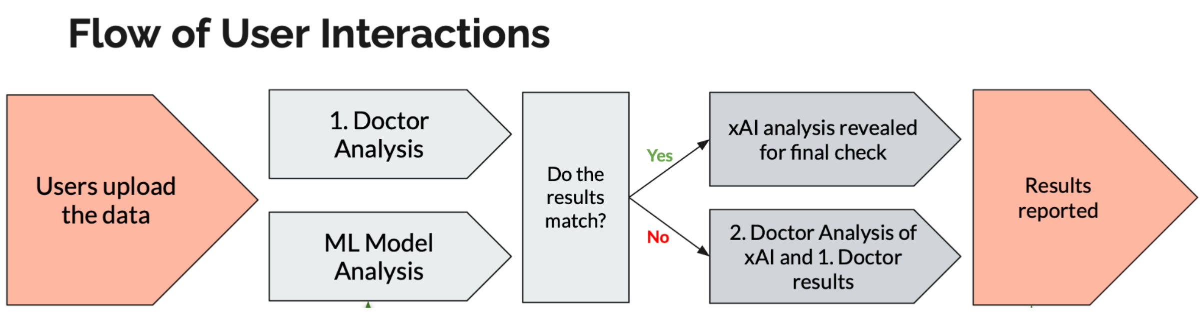

Fig. 5: Flow of user interactions.

One other aspect is about the applicability and accuracy of the XAI models usage in the mammography domain. Firstly, the concept of explainability and its taxonomy depicted in Figure 6 has to be understood. It is possible to implement both ad-hoc (inherently explainable model) and post-hoc (model is “deciphered” after the classification using additional methods). Our application is planned to be global (able to explain all types of classifications), in-model (explainability methods are integrated into the model itself), or post-model, visualized explainable mode [25]. According to our research on implementing these models, there are both attribution-based and non-attribution-based models. For the attribution-based systems that are designed to provide attributions for model outputs, the Integrated Gradients attribution method and the SmoothGrad noise reduction algorithm are used to visualize and annotate the features of a CNN for breast images [26]. Additionally, for mass classification of mammograms, a group used two different CNN models, AlexNet and GoogleNet, with the saliency maps to visualize the interest areas with an accuracy of 0.89 and 0.92, respectively [27]. For the non-attribution attention-based models, MDNet was used to map between medical images and corresponding diagnostic reports. With an image model and a language model combined, the attention mechanisms are used to visualize the classification (they got 78.4 diagnostic conclusion accuracy) [28]. Moreover, SAUNet (interpretable version of UNet) added a parallel secondary shape stream to capture important shape-based information with the regular texture features of the images. The architecture used an attention module in the decoder part of the U-Net. The spatial and shape attention maps were generated using SmoothGrad to visualize the high activation region of the images [29]. They got an accuracy between 0.88 and 0.93. Furthermore, as an example of text justification models, a model that takes the visual features of a classifier and the embedding of predictions was used to create sentences about diagnosis and visual heatmaps for breast mass classification. They developed a gated recurrent unit-based recurrent neural network with hierarchical attention for mortality prediction and achieving AUROC score 0.86 [30].

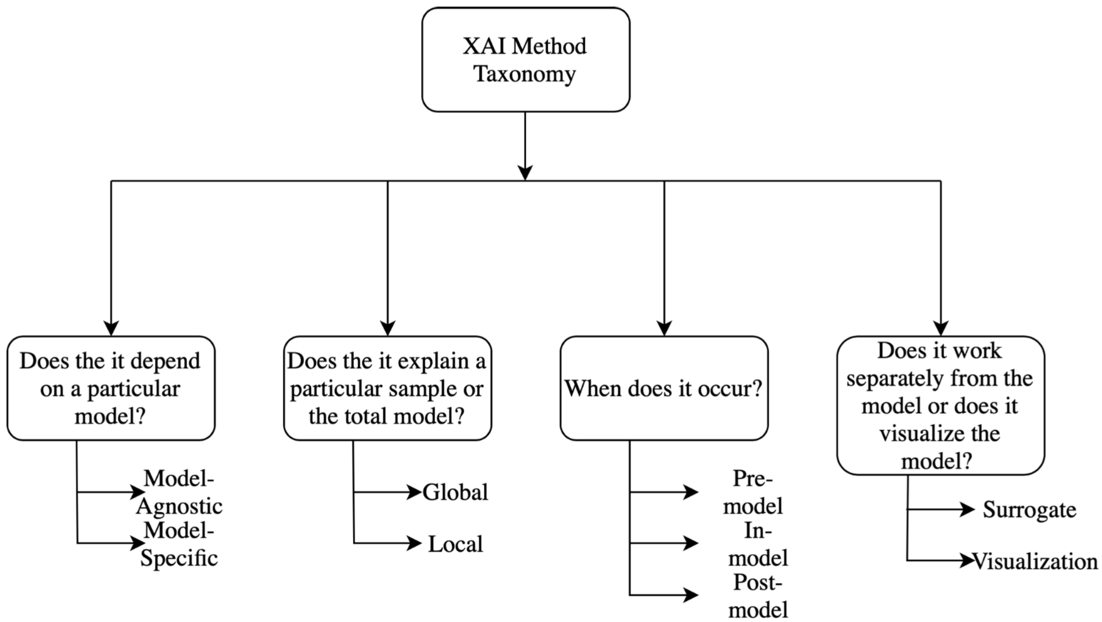

Fig. 6: Taxonomy of XAI Methods

4. Glossary

Term	Definition
XAI	Explainable Artificial Intelligence
HIPAA	Health Insurance Portability and Accountability Act
KETEM	Cancer Early Diagnosis, Screening and Education Center
BI-RADS	Breast Imaging-Reporting and Data System

5. References

[1] Breast cancer facts and statistics 2023. Available at: https://www.breastcancer.org/facts-statistics (Accessed: 28 October 2023).

[2] Prof.Dr.M. (2018) Meme Kanseri belirtileri, Teşhisi Ve Tedavi yöntemleri: Anadolu Sağlık Merkezi, Anadolu Sağlık. Available at: https://www.anadolusaglik.org/blog/meme-kanseri (Accessed: 28 October 2023).

[3] Harbeck, N. *et al.* (2019) 'Breast cancer', *Nature Reviews Disease Primers*, 5(1). doi:10.1038/s41572-019-0111-2.

[4] KPSPublic Web Service. Available at: https://tckimlik.nvi.gov.tr/Service/KPSPublic.asmx?op=TCKimlikNoDogrula (Accessed: 27 October 2023).

[5] General Data Protection Regulation. Available at: https://gdpr-info.eu/ (Accessed: 27 October 2023).

[6] Kişisel Verileri Koruma Kanunu. Available at: https://www.kvkk.gov.tr/ (Accessed: 27 October 2023).

[7] e-Nabız| T.C. Sağlık Bakanlığı. Available at: https://enabiz.gov.tr/ (Accessed: 27 October 2023).

[8] *Turkey breast cancer therapeutics market analysis - INSIGHTS10*. Available at: https://www.insights10.com/report/turkey-breast-cancer-therapeutics-market-analysis/ (Accessed: 28 October 2023).

[9] *Toplumsal CİNSİYET istatistikleri gender statistics 2022*. Available at: https://www.tuik.gov.tr/media/announcements/toplumsal_cinsiyet_istatistikleri.pdf (Accessed: 28 October 2023).

[10] *Doktora Sor, Danış, Tıbbi ikinci Görüş Al. Hastalar Soruyor*. Available at: https://hastalarsoruyor.com/ (Accessed: 28 October 2023).

[11] *Remote Second Opinion, Online Medical Archive Medicopin*. Available at: https://www.medicopin.com/pages/1202/tr-TR/medicopin (Accessed: 28 October 2023).

[12] *Smart breast ai: Transpara (R) from Screenpoint Medical* (2023) *Screen Point*. Available at: https://screenpoint-medical.com/ (Accessed: 31 October 2023).

[13] *Artificial Intelligence, the Hesperus of Healthcare: ICterra Information and Communication Technologies* (2023) *ICterra*. Available at: https://www.icterra.com/artificial-intelligence-the-hesperus-of-healthcare/ (Accessed: 31 October 2023).

[14] *Syngo.via Ile Yapay zeka Ile Entegre Mamografi Teknolojileri syngo.via ile Yapay Zeka ile Entegre Mamografi Teknolojileri - Siemens Healthineers Türkiye*. Available at: https://www.siemens-healthineers.com/tr/mammography/reading-reporting/syngo-breastcare#05270776 (Accessed: 31 October 2023).

[15] The Cancer Imaging Archive, https://dev.cancerimagingarchive.net/ (Accessed: 28 October 2023).

[16] A. S. Alsolami et al., "King Abdulaziz University Breast Cancer Mammogram Dataset (Kau-BCMD)," MDPI, https://www.mdpi.com/2306-5729/6/11/111 (Accessed: 28 October 2023).

[17] "Explore Image Data," IDC, https://portal.imaging.datacommons.cancer.gov/explore/ (Accessed: 28 October 2023).

[18] "OMI-DB Database Information," NHS choices, https://medphys.royalsurrey.nhs.uk/omidb/stats_table/ (Accessed: 28 October 2023).

[19] S. M. McKinney et al., "International Evaluation of an AI system for breast cancer screening," Nature, vol. 577, no. 7788, pp. 89–94, 2020. doi:10.1038/s41586-019-1799-6

[20] S. A. Agnes, J. Anitha, S. I. Pandian, and J. D. Peter, "Classification of mammogram images using multiscale all convolutional neural network (MA-CNN)," *Journal of Medical Systems*, vol. 44, no. 1, 2019. doi:10.1007/s10916-019-1494-z

[21] P. Bharati and A. Pramanik, "Deep learning techniques—R-CNN to mask R-CNN: A survey," *Computational Intelligence in Pattern Recognition*, pp. 657–668, 2019. doi:10.1007/978-981-13-9042-5_56

[22] Gaube, S. *et al.* (2021) 'Do as AI say: Susceptibility in deployment of clinical decision-AIDS', *npj Digital Medicine*, 4(1). doi:10.1038/s41746-021-00385-9.

[23] Brown, J., Bryan, S. and Warren, R. (1996) 'Mammography screening: An incremental cost effectiveness analysis of double versus single reading of mammograms', *BMJ*, 312(7034), pp. 809–812. doi:10.1136/bmj.312.7034.809.

[24] Gulum, M.A., Trombley, C.M. and Kantardzic, M. (2021) 'A review of Explainable Deep Learning Cancer Detection Models in medical imaging', *Applied Sciences*, 11(10), p. 4573. doi:10.3390/app11104573.

[25] Singh, A., Sengupta, S. and Lakshminarayanan, V. (2020) *Explainable deep learning models in medical image analysis*, *arXiv.org*. Available at: https://arxiv.org/abs/2005.13799 (Accessed: 31 October 2023).

[26] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.e.a. Going deeper with convolutions. Proc. IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[27] MD, S.Y. (no date) *Interpretable predictions of clinical outcomes with an attention-based recurrent neural network*, *ACM-BCB ... ... : the ... ACM Conference on Bioinformatics, Computational Biology and Biomedicine. ACM Conference on Bioinformatics, Computational Biology and Biomedicine*. Available at: https://pubmed.ncbi.nlm.nih.gov/32577628/ (Accessed: 31 October 2023).

[28] Zhang, Z.; Xie, Y.; Xing, F.; McGough, M.; Yang, L. Mdnet: A semantically and visually interpretable medical image diagnosis network. Proc. IEEE conference on computer vision and pattern recognition, 2017, pp.6428–6436.

[29] Sun, J.; Darbeha, F.; Zaidi, M.; Wang, B. SAUNet: Shape Attentive U-Net for Interpretable Medical Image
Segmentation. *arXiv preprint arXiv:2001.07645* **2020**.

[30] Sha, Y. and Wang, M.D. (2017) 'Interpretable predictions of clinical outcomes with an attention-based recurrent neural network', *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics* [Preprint]. doi:10.1145/3107411.3107445.