Research Protocol:

A Study on Artificial Intelligence Algorithms for Breast Cancer Classification from Histopathology Images

Principal Investigator:

Dr. Taufiq Hasan
Associate Professor, Biomedical Engineering,
Bangladesh University of Engineering and Technology (BUET), Dhaka – 1205.

Co-Principal Investigator:

Dr. Farida Arjuman
MBBS(CMC), BCS(Health),
FCPS (Histopathology),MCPS(Clinical Pathology)
Associate Professor & Head
Department of Histopathology
National Institute of Cancer Research & Hospital (NICRH)
Mohakhali, Dhaka-1212

Date of submission:
18.03.2024

Application for Ethical Clearance

National Institute of Cancer Research and Hospital
Mohakhali, Dhaka-1212

1. **Principal Investigator(s):**
Name: Dr. Taufiq Hasan
Qualification: B.Sc, M.Sc, PhD (UT Dallas, TX, USA)
Detailed address: Department of Biomedical Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka-1205.
Mobile: 01817579844 Telephone (Off./Res.): N/A
e-mail: taufiq@bme.buet.ac.bd

2. **Co-Principal Investigator(s):**
Name: Dr. Farida Arjuman
Qualification: MBBS(CMC), BCS(Health), FCPS(Histopathology), MCPS(Clinical Pathology)
Detailed address: Department of Histopathology, National Institute of Cancer Research & Hospital (NICRH), Mohakhali, Dhaka-1212
Mobile: 01715520165 Telephone (Off./Res.): N/A
e-mail: drarju35cmc@gmail.com

3. **Place of Study/Institution(s):**

   i) National Institute of Cancer Research & Hospital (NICRH), Mohakhali, Dhaka-1212

4. **Title of Study:** A Study on Artificial Intelligence Algorithms for Breast Cancer Classification from Histopathology Images
5. **Type of Study:** Prospective and retrospective case control study
6. **Duration of study:** 1 year
7. **Total cost:** N/A
8. **Funding Agency:** N/A

I agree to obtain approval of the NICRH Ethical Review Committee for any changes involving the rights and welfare of subjects or any changes of the Methodology before making any such changes.

Dr. Taufiq Hasan
Associate Professor, Department of Biomedical Engineering,
Bangladesh University of Engineering and Technology (BUET), Dhaka – 1205.

Dr. Farida Arjuman, MBBS(CMC), BCS(Health), FCPS (Histopathology),MCPS(Clinical Pathology)
Associate Professor & Head, Department of Histopathology
National Institute of Cancer Research & Hospital (NICRH), Mohakhali, Dhaka-1212

Circle the appropriate answer to each of the following
(If not Applicable write N/A)

1. Source of Population:

a. ILL PArticipant (Yes) No

b. Non ILL Participant (Yes) No

c. Minors or persons under guardianship (Yes) No

2. Does the study involve?

a. Physical risks to the subjects Yes (No)

b. Social Risks Yes (No)

c. Psychological risks to subjects Yes (No)

d. Discomfort to subjects Yes (No)

e. Invasion of the body Yes (No)

f. Invasion of privacy (Yes) No

g. Disclosure of information damaging to subject or others (Yes) No

3. Does the study involve?

a. Use of records (Hospital, medical, Death, birth or other) (Yes) No

b. Use of foetal tissue or abortus Yes (No)

c. Use of organs or body fluids Yes (No)

4. Are subjects clearly informed about?

a. Nature and purposes of study (Yes) No

b. Procedures to be followed including alternatives used N/A

c. Physical risks N/A

d. Private questions (Yes) No

e. Invasion of the body N/A

f. Benefits to be derived (Yes) No

g. Right to refuse to participate or to withdraw from study (Yes) No

h. Confidential handling of data (Yes) No

i. Compensation where there are risks or loss of working time or privacy is involved in any particular procedure (Yes) No

5. Will signed consent form/verbal consent be required?

a. From subjects (Yes) No

b. From parent or guardian (if subjects are minors) (Yes) No

6. Will precautions be taken to protect anonymity of subject? (Yes) No

RESEARCH PROTOCOL

- **Project Title:** A Study on Artificial Intelligence Algorithms for Breast Cancer Classification from Histopathology Images

- **Summary:** Breast cancer, a widespread and potentially fatal illness, emphasizes the urgent requirement for early and precise detection to enhance patient outcomes. The current diagnostic framework, characterized by manual histopathological examination, exhibits inherent drawbacks such as subjectivity and limited throughput. In response, our research aims to transform breast cancer diagnostics by leveraging advanced computational techniques, particularly deep learning. Breast cancer can be classified into two main subtypes: invasive and noninvasive.

  Noninvasive breast cancer, often termed in situ breast cancer, is characterized by abnormal cells confined within the milk ducts (ductal carcinoma in situ or DCIS) or lobules (lobular carcinoma in situ or LCIS) without invading surrounding tissues. This type is considered an early stage and typically not life-threatening on its own. Invasive breast cancers are those that spread from the original site (either the milk ducts or the lobules) into the surrounding breast tissue. These constitute approximately 70% of all breast cancer cases and generally have a poorer prognosis compared to the in-situ subtypes. [1]

  Medical imaging of breasts can be acquired through various techniques, such as MRI scans, mammography, ultrasound, thermography, computed tomography scans, and histopathology. Among these approaches, the histopathology test serves as the gold standard for the clinical diagnosis of cancer. [2]

  The proposed AI model aims to streamline and enhance the analysis of histopathological images for the classification of invasive and noninvasive breast cancer, surpassing conventional methods and providing a reliable means of identifying cancerous regions. This promises to significantly improve the accuracy and efficiency of breast cancer diagnostics, meeting the urgent need for dependable and scalable solutions. By automating the complex process of breast cancer histopathological image analysis, our goal is to democratize screening, making it more accessible and reaching underserved populations. Moreover, our model goes beyond technological innovation; it addresses broader issues of accessibility and scalability, particularly in low-income settings. The research's focus on domain adaptation is crucial, ensuring the model's accuracy and reliability across various health facilities. This involves accommodating differences in resources, equipment, and demographic factors, making it a versatile and adaptable solution for diverse contexts.

  In summary, our research not only aims to advance the technological frontier in breast cancer diagnostics but also seeks to drive a transformative change in accessibility,

efficiency, and reliability. By developing an AI model that tackles the specific challenges of traditional methods, we aim to make a meaningful impact on breast cancer screening, ultimately leading to early detection, tailored treatment, and improved outcomes for patients worldwide.

- **Introduction:** Breast cancer is a condition where abnormal breast cells proliferate uncontrollably, leading to the formation of tumors. If left untreated, these tumors can metastasize, spreading to other parts of the body and potentially becoming fatal. Breast cancer originates within the milk ducts or lobules of the breast. The initial form, known as in situ, is non-life-threatening. However, cancer cells can infiltrate nearby breast tissue, resulting in the development of tumors that manifest as lumps or thickening. Invasive breast cancers have the potential to metastasize to nearby lymph nodes or other organs, posing a significant risk of fatality.

  Breast cancer comprises 12.5% of all new cancer cases annually worldwide, rendering it the most prevalent cancer globally. In 2020 alone, 2.3 million women received a breast cancer diagnosis, with approximately 685,000 succumbing to the disease, solidifying its status as the most common cancer. By the close of 2020, 7.8 million women had survived a breast cancer diagnosis within the previous five years, establishing it as the world's foremost prevalent cancer. Breast cancer afflicts women in every nation post-puberty, with incidence rates escalating in later stages of life. [3]

  Breast cancer detection typically involves palpation, followed by non-invasive mammogram-based identification of critical points. For confirmation and comprehensive disease profiling, invasive microscopic examination of biopsy samples extracted from the breast's critical mass is conducted. [4]

  Early and precise detection of breast cancer is paramount in reducing mortality rates, significantly improving the prognosis and elevating patient survival rates up to 50%. Treatment approaches vary depending on individual factors, cancer type, and extent of spread, often combining surgical intervention, radiation therapy, and medication. In histopathology, the standard procedure for diagnosing breast cancer involves examining a hematoxylin and eosin (H&E) stained tissue biopsy under a microscope. However, this method relies on manual, qualitative inspection, leading to potential investigator bias and low throughput. [5]

  Manual examination of histopathological images is labor-intensive and taxing. Moreover, there's a significant risk of overlooking small cancerous regions on slides, especially as malignant cells may be sparsely distributed. In this scenario, machine learning-based automated inspection techniques offer rapid and dependable identification of invasive and noninvasive cancerous cells and regions on the slide. [5] Diagnoses heavily depend on the expertise of medical professionals, introducing subjectivity and limited repeatability, with documented inaccuracies even among experienced practitioners.

Traditional vision screening of histopathological images via light microscopy poses risks of misdiagnosis and struggles with accessibility due to time and personnel limitations. The primary focus is on leveraging advanced computational techniques to automate analysis, addressing inherent drawbacks in traditional screening methods. Through AI approaches, our aim is to create a model surpassing the accuracy of visual assessments, providing a reliable and efficient means of identifying transformation zones and detecting early signs of breast abnormalities. Central to our objective is the recognition of the need for scalable and cost-effective solutions, particularly in low-income settings. The proposed AI model seeks to mitigate the shortage of trained personnel by automating breast cancer histopathological image analysis, streamlining screening and enhancing accessibility. Moreover, we aim to incorporate domain adaptation techniques to ensure the model's precision and resilience across diverse health facilities, considering challenges such as resource constraints, equipment variations, and demographic diversity.

- **Objectives:** Our primary goal is to compile a diverse dataset of breast tissue histopathology images, incorporating samples from various sources to ensure representation across patient demographics and imaging equipment. Using this dataset, we aim to develop a robust AI model for automated classification of invasive and noninvasive breast cancer within histopathology images. Specific objectives include:

  a) Clinical Validation: Evaluate the AI model's performance through comprehensive clinical validation, measuring sensitivity, specificity, and accuracy in classifying invasive and noninvasive breast cancer types in histopathology images.

  b) Usability Assessment: Assess the model's usability by involving healthcare professionals in the evaluation process, measuring integration into clinical workflows, result interpretability, and time efficiency during histopathology examinations.

  c) Impact on Early Detection: Investigate the AI model's impact on early breast cancer detection, analyzing its contribution to identifying cancer types at earlier stages for potentially improved patient outcomes.

  d) Generalizability: Evaluate the model's generalizability across diverse datasets, including local sources, assessing its performance in different regions and healthcare settings for broader clinical applicability.

  e) Impact on Patient Outcomes: Examine the AI model's impact on patient outcomes, such as reduced diagnosis time, increased accuracy in cancer type identification, and enhanced treatment planning, considering potential benefits to patient well-being and satisfaction.

  f) Clinical Decision Support: Evaluate the model's effectiveness as a clinical decision support tool, measuring its contribution to accurate diagnoses and treatment plans during histopathology examinations.

Through these objectives, our research aims to advance breast cancer diagnosis, providing a comprehensive solution that integrates seamlessly with clinical practices and positively influences patient outcomes.

- **Rationale:** Histopathology images contain valuable information about tissue structure, but manual interpretation can be subjective and time-intensive. Deep learning offers a promising solution by automatically learning complex patterns, potentially improving the accuracy and efficiency of breast cancer classification. This research aims to leverage deep neural networks to analyze extensive datasets of histopathology images, distinguishing between invasive and noninvasive breast cancer with high accuracy. Automating this classification process aims to reduce human error, expedite diagnoses, and support personalized treatment approaches. Ultimately, the goal is to refine breast cancer diagnosis, enabling earlier intervention and better patient outcomes.

- **Methodology:** In the clinic-based approach for histopathological breast cancer screening data collection, patient information will be directly gathered within healthcare facilities to develop an AI algorithm for automated cancer detection. This involves capturing essential metadata and clinical details, including histopathological breast images. Utilizing clinic resources, the collected data will train the AI algorithm, offering insights into abnormal findings likelihood during screening. Notably, privacy is prioritized, excluding personally identifiable information in the data collection process.

  a) **Study Population**: This investigation comprises a prospective and retrospective case-control study designed to develop and assess an AI-based diagnostic tool for breast cancer detection through visual examination. The patient cohorts will include confirmed breast cancer cases diagnosed via gold standard tests and individuals with breast abnormalities resembling cancer (such as Adenosis, Duct Ectasia, Ductal or Lobular Hyperplasia, Fibroadenomas, Fibrosis, Mastitis, Simple Cysts, or other precancerous lesions), along with corresponding confirmatory test results. Additionally, data from individuals with suspected breast cancer (test status unknown) or normal breast condition will be included to support unsupervised learning. The primary aim is to assess the algorithm's ability to differentiate between these patient categories through analysis of their medical image data.

  - **Inclusion criteria:**

- Female patients of age 18 years or older can be selected as subjects.
- Individuals willing to participate in breast cancer screening.
- Availability for biopsy examination.
- Women with no current or prior diagnosis of breast cancer.
- Availability of relevant medical records for confirmation and comparison purposes.

- **Exclusion criteria:**

- Pregnant women are excluded due to potential impacts on screening results and the necessity for special considerations during pregnancy.
- Individuals with severe medical conditions or circumstances that may render histopathologic examination inappropriate or unsafe are excluded.
- Patients with conditions that could interfere with the accuracy of screening results are excluded.
- Follow-up screenings are not included in this study.

b) **Sample size calculation**: Following the methodology of sample size calculation to evaluate diagnostic performance (for algorithms in this case), we use the following method of calculation for a target sensitivity and specificity incorporating disease prevalence [7]:

$$N_{sen} \geq \frac{Z^2(1-\alpha)/2 \times sen\,(1-sen)}{c^2 \times p}$$

$$N_{spec} \geq \frac{Z^2(1-\alpha)/2 \times spec\,(1-spec)}{c^2 \times (1-p)}$$

N_{sen} = Required sample size for sensitivity, *sen*
N_{spec} = Required sample size for specificity, *spec*
α = Confidence level
$Z^2(1-\alpha)/2$ = The *Z*-score for the confidence level α
p = Prevalence of breast cancer
c = Confidence interval
sen = Estimated sensitivity
spec = Estimated specificity

- The final sample size *N* is calculated based on the larger of the two values N_{sen} and N_{spec}. In this study, as we are using deep learning techniques, we need the medical image samples to reflect the various cancerous, precancerous and normal cases. For our analysis, we assume an 80% confidence level (α = 0.80) and prevalence of breast cancer for self-reported instances, p = 0.261 [23]. A confidence interval of 5% is selected (c = 0.05). We estimate the sensitivity and specificity of our algorithms to be at 75% (*sen* = 0.75 and *spec* = 0.75). Using these values, we obtain the sample size values: N_{sen} = 471 and N_{spec} = 166. Since $N_{sen} > N_{spec}$, the final sample size for the study will be:

$$N = N_{sen} = 471$$

We can set a target of $N = 500$.

c) **Procedures:** During the prospective phase of the study, data collection will occur in clinical settings. Eligible patients will be approached during clinic visits and presented with a consent form for their review. Upon agreement, each participant will be assigned a unique identifier to maintain data integrity and ensure patient anonymity. Subsequently, patients will provide essential clinical information. Healthcare professionals will conduct histopathological examinations, and specialized equipment available in the clinic will be used to capture images of the breast region. This method guarantees the collection of non-identifiable data, unique data identifiers, and informed consent directly within the clinical setting, meeting the specific requirements of histopathological breast cancer screening.

In the retrospective phase of the study, additional data will be retrieved from the clinic's database of prior screenings. Due to the retrospective nature, obtaining consent from individual patients is impractical. To address privacy concerns, any personally identifiable patient information will be systematically excluded from the collected data. This ensures compliance with privacy standards while leveraging valuable information from past screenings. While informed consent cannot be obtained for this dataset, the study will strictly adhere to ethical guidelines and prioritize patient privacy during the data extraction process from the clinic's database.

d) **Methods of Data Collection**:

- Patient ID (unique identification number)
- Date & Time of data input
- Age
- Histopathological breast image data
- Region of Interest (ROI) Annotations:
  - Lesion size
  - Provisional diagnosis
  - Suggested management (e.g., routine screening, biopsy, etc.)
  - Patient’s relevant medical/surgical history
- Additional Information (if available):
- Relevant diagnostic test results
- Reproductive history

The hematoxylin and eosin (H&E) stained images obtained from histopathological assessment, along with relevant metadata, will significantly contribute to algorithm design and image analysis. These

images' standard test status and corresponding results will serve as the ground truth for our detection algorithm.

e) **Data Interpretation**: Following the completion of data collection, exploratory data analysis will be conducted to assess patient distribution across various parameters, including transformation zone type, histopathology, and relevant clinical factors. Subsequently, a balanced dataset will be curated to ensure fair representation of image samples.

This curated dataset will serve as the foundation for training and evaluating the artificial intelligence (AI) algorithm developed for automatic classification and detection of abnormalities in histopathological breast screening cases. The trained model will generate predictions regarding the presence or absence of abnormalities in the breast images.

To validate the accuracy and clinical relevance of the algorithm's predictions, the results will undergo verification and analysis in collaboration with expert histopathologists. Their insights will be invaluable in refining the automatic algorithm, thereby improving its precision in classifying and detecting potential abnormalities during histopathological breast screenings.

f) **Statistical Analysis**: Once the dataset becomes available, statistical analysis will center on evaluating the performance of the developed artificial intelligence algorithms for automatic classification and detection of abnormalities in histopathological breast screening cases. Key metrics such as sensitivity, specificity, precision, recall, F-1 score, and overall accuracy will be utilized for rigorous evaluation. To ensure robustness, experiments involving n-fold cross-validation will be performed, with mean values and standard deviations of performance metrics calculated. This analysis will conclude with a comparative assessment of various deep learning models, assessing their effectiveness in the automatic classification task.

Utilization of Results: The study results will be pivotal in the development and assessment of an automatic detection algorithm for breast cancer screening, with the following primary objectives:

a) Automatic Detection: Evaluate and detect potential abnormalities in histopathological images, furnishing an automated screening tool for efficient identification of breast cancer risk.

b) Clinical Decision Support: Aid healthcare professionals by furnishing an automated analysis to assist in prioritizing treatment strategies and making informed clinical decisions based on the algorithm's findings.

c) Systematic Data Storage: Enable systematic storage of histopathological images and pertinent metadata to bolster ongoing research efforts, all while safeguarding individuals' privacy during the automated detection process.

These outcomes are poised to significantly enhance the efficiency and accuracy of breast cancer screening, offering valuable support to healthcare professionals in Bangladesh.

- **Facilities:** In the case of collecting data from histopathological breast cancer screening, no additional facilities are required except those the clinics are already equipped with (e.g. image acquisition and data storage capacity).

- **Approval / Forwarding of the Head of Department / Institute / IRB.** A forwarding letter from the Department of Biomedical Engineering, Bangladesh University of Engineering and Technology (BUET) is attached.
- **Flow Chart:** The data collection activity and the corresponding timeline is described in the Gantt chart below:

No.	Work/Activities	Months					
		1	2	3	4	5	6
1	Initial algorithm with publicly available data						
2	Collection of images and metadata						
3	Homogenization and organisation of data						
4	Data processing and analysis						

- **Ethical Implications:** The proposed retrospective/prospective case-control study based on medical records is designed to develop and evaluate the diagnostic performance of Artificial Intelligence (AI) algorithms that will detect breast cancer stage from histopathological images. The study poses minimal risk to the subjects according to the HHS (USA) definition.

- **References:**

1. Cruz-Roa, Á., Gilmore, H., Basavanhally, A., Feldman, M. D., Ganesan, S., Shih, N., Tomaszewski, J. E., González, F. A., & Madabhushi, A. (2017). Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Scientific Reports*, *7*(1). https://doi.org/10.1038/srep46450

2. Toma, T., Biswas, S., Miah, M. S., Alibakhshikenari, M., Virdee, B. S., Fernando, S., Rahman, H., Ali, S. M., Arpanaei, F., Hossain, M. A., Rahman, M. M., Niu, M., Parchin, N. O., & Livreri, P. (2023). Breast cancer detection based on simplified deep learning technique with histopathological image using BREAKHIS database. *Radio Science*, *58*(11). https://doi.org/10.1029/2023rs007761

3. World Health Organization: WHO & World Health Organization: WHO. (2023, July 12). *Breast cancer*. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

4. Majeed, H., Kandel, M. E., Han, K., Luo, Z., Macias, V., Tangella, K., Balla, A., & Popescu, G. (2015). Breast cancer diagnosis using spatial light interference microscopy. *Journal of Biomedical Optics*, *20*(11), 111210. https://doi.org/10.1117/1.jbo.20.11.111210
5. Das, K., Conjeti, S., Chatterjee, J., & Sheet, D. (2020). Detection of breast cancer from whole slide histopathological images using deep multiple instance CNN. *IEEE Access*, *8*, 213502–213511. https://doi.org/10.1109/access.2020.3040106

6. LabMedica International. (2015, September 16). Breast cancer diagnosis uses spatial light interference microscopy. *mobile.Labmedica.com*. https://mobile.labmedica.com/?option=com_mobile_article&Itemid=294760436
7. Hoque, M. R., Haque, E., & Karim, M. R. (2020). Cervical cancer in low‐income countries: A Bangladeshi perspective. *International Journal of Gynecology & Obstetrics*, *152*(1), 19–25. https://doi.org/10.1002/ijgo.13400

INFORMED CONSENT FORM (PROSPECTIVE STUDY)

Title: A Study on Artificial Intelligence Algorithms for Breast Cancer Classification from Histopathology Images

Principal Investigator:

Dr. Taufiq Hasan, Associate Professor,
Department of Biomedical Engineering, BUET, Dhaka – 1205

Please check all that are applicable:

☐ I am an adult participant in this study.

☐ I am the parent or guardian granting permission for a child in this study (the use of "you" refers to "your child" or “your ward.”)

Are you participating in any other research studies? (Yes/No) _____

PURPOSE OF PROJECT

You are invited to participate in a project entitled **Clinical Evaluation of Artificial Intelligence Algorithms for Breast Cancer Detection from Histopathology Images.** This project focuses on developing an algorithm that will automatically detect invasive and non-invasive cancer types using histopathological images.The period of the study will be between February 10, 2024 to August 31, 2024. You are selected as a possible participant in the study

If you decide to terminate your participation in this study, you should notify Dr. Taufiq Hasan at +8801817579844.

PROCEDURES

The procedure for collecting histopathology slide images for a breast cancer detection typing AI algorithm involves recruiting eligible female patients, obtaining informed consent, conducting routine clinical visits with breast tissue biopsies, preparing histopathology slides, capturing digital images, encrypting and anonymizing data, extracting retrospective information, compiling a diverse dataset, and ensuring ongoing privacy and ethical oversight. The goal is to develop an accurate AI algorithm while prioritising patient privacy and adhering to ethical guidelines.

The data collected during these sessions will be put into one of the databases of the Department of BME, BUET and will be used for future research.

No traditionally used identifying information about you, such as your name, address, telephone number, or national ID number, will be put into the public database. Your privacy is very important to us, and we will use safety measures to protect it. Despite all the safety measures that we will use, we cannot guarantee that your identity will never become known

PARTICIPANT RESPONSIBILITIES

As a participant, your responsibilities include:

- Adhere to instructions provided by healthcare professionals during clinic visits for data collection.
- Contact the data collection coordinator or research study staff promptly if you have any questions or concerns.
- If you are a healthcare professional, particularly a histopathologist, ensure accurate maintenance of the image ID to prevent data mix-ups.
- Notify the coordinator or research staff directly if you decide to withdraw from the study on the clinic-based system, ensuring seamless communication with the study staff.

WITHDRAWAL FROM STUDY

If you first agree to participate and then you change your mind, you are free to withdraw your consent and discontinue your participation at any time.

If you decide to withdraw your consent to participate in this study, you can directly inform Dr. Taufiq Hasan at +8801817579844 or notify the study staff.

If you withdraw from the study for any reason all the data acquired from you will be removed from the databases.

YOU MUST NOTIFY ABOUT YOUR DISSENT WITHIN THE PERIOD OF THE STUDY BEING CONDUCTED.

The Protocol Director may also withdraw you from the study without your consent for one or more of the following reasons:

- Failure to follow the instructions during data collection.
- The Protocol Director decides that continuing your participation could be harmful to you.
- You need other treatments that are not allowed in the study.
- The study is cancelled.
- Unanticipated circumstances.
- Other administrative reasons

POSSIBLE RISKS, DISCOMFORTS, AND INCONVENIENCES

There may be mild discomfort during the procedure and, in rare cases, minimal bleeding or infection at the biopsy site. You should contact the Protocol Director or the study staff if you have any questions.

POTENTIAL BENEFITS

We cannot and do not guarantee or promise that you will receive any benefits from this study. The study is being conducted for humanitarian purposes, which envision a long-term goal that will benefit patients suffering from or at the risk of having breast cancer.

PARTICIPANT'S RIGHTS

You should not feel obligated to agree to participate. Your questions should be answered clearly and to your satisfaction. If you decide not to participate, notify the project coordinator or the study staff.

CONFIDENTIALITY

The results of this research study may be presented at scientific or medical meetings or published in scientific journals. Your identity and/or your personal health information will not be disclosed except as authorized by you or as required by law.

CONTACT INFORMATION

Questions, Concerns, or Complaints: If you have any questions, concerns or complaints about this research study, its procedures, risks and benefits, or alternative courses of treatment, you should ask the data collection coordinator, Dr. Taufiq Hasan. You should also contact him at any time if you feel you have been harmed by being a part of this study at +8801817579844.

May we contact you about future studies that may be of your interest? (Yes/No)

Checking on the box with a click means you agree to be in this study.

☐ Agreed

JUSTIFICATIONS FOR INFORMED CONSENT WAIVER (RETROSPECTIVE STUDY)

The proposed research is a case-control study that only uses medical records of histopathology slides of breast tissues. As the number of samples required is large (500 according to our sample size calculation) compared to the monthly recorded cases, it is not possible to return to the patients for acquiring their consent in case of previously recorded data. Considering the high potential impact of the study, we earnestly request waiver of informed consent. The justifications for this request is provided below:

1. On the topic of informed consent waiver, the regulatory authority of the United States, Department of Health and Human Services (HHS) mentions in 45 C.F.R. § 46.116(d):

> *"An IRB may approve a consent procedure which does not include, or which alters, some or all of the elements of informed consent set forth in this section, or waive the requirements to obtain informed consent provided the IRB finds and documents that:*
>
> *(1) The research involves no more than minimal risk to the subjects;*
>
> *(2) The waiver or alteration will not adversely affect the rights and welfare of the subjects;*
>
> *(3) The research could not practicably be carried out without the waiver or alteration; …"*

2. On the topic of case control studies, the Bangladesh Medical Research Council (BMRC) "Ethical Guidelines for Conducting Research Studies Involving Human Subjects" mentions on Page 45:

> *"However, if it entails only a review of medical records, informed consent may not be required and indeed may very often not be feasible".*

In light of the abovementioned regulatory guidelines described by HHS (USA) and BMRC, we want to highlight that our study fulfils all of the above conditions as explained below:

1. The proposed research involves no more than minimal risk to subjects. According to HHS document 45 C.F.R. § 46.102(i):

> *"Minimal risk means that the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests."*

We will use de-identified medical records for the study. This falls into the category of minimal risk as described above.

2. The research could not be carried out practicably without the waiver or alteration. Training artificial intelligence algorithms require a huge amount of data from a large number of

patients due to the overfitting problem (further detailed in the sample size calculation section below). It is not practically possible to collect signed or verbal informed consent from each patient from each hospital. In addition, most of the patients hospitalized due to COVID-19 are in severe conditions and are not capable of providing consent.

3. The waiver or alteration will not adversely affect the rights and welfare of the subjects. Only their digital medical images and clinical history data will be used for the research. Personal identifying information (name, etc.) will be removed at the source.

With the explanations provided above, we hope that the ethical review committee will be kind enough to approve our request to waive the informed consent requirement for this case-control study using medical records. As we have mentioned above, the study is of potentially high impact and unfortunately not feasible without the waiver.