Artificial Intelligence-Based Breast and Cervical Cancer Diagnosis and Management System

Elbetel Taye Zewde[1,2], Mizanu Zelalem Degu[2,3], and Gizeaddis Lamesgin Simegn[1,2](✉)

[1] Biomedical Imaging Unit, School of Biomedical Engineering, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia
gizeaddis.lamesgin@ju.edu.et
[2] AI and Biomedical Imaging Research Unit, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia
[3] Faculty of Computing and Informatics, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia

Abstract. Breast cancer and cervical cancer are two of the most common and deadly malignancies in women. Early diagnosis and treatment can save lives and improve quality of life. However, there is a shortage of pathologists and physicians in most developing countries, including Ethiopia, preventing many breast and cervical cancer patients from early cancer screening. Many women, particularly in low resource settings, have limited access to early diagnosis of breast and cervical cancer and receive poor treatment which in turn increases the morbidity and mortality due to these cancers. In this paper, an integrated intelligent decision support system is proposed for the diagnosis and management of breast and cervical cancer using multimodal im-age data. The system includes breast cancer type, sub-type and grade classification, cervix type (transformation zone) detection and classification, pap smear image classification, and histopathology-based cervical cancer type classification. In addition, patient registration, data retrieval, and storage as well as cancer statistical analysis mechanisms are integrated into the proposed system. A ResNet152 deep learning model was used for classification tasks and satisfactory results were achieved when testing the model. The developed system was deployed to an offline web page which has added the advantage of storing the digital medical images and the labeled results for future use by the physicians or other researchers.

Keywords: Breast cancer · Cervical cancer · Decision support system · Screening · Histopathological images · Cancer management

1 Introduction

The world health organization (WHO) estimated that there were 9.6 million cancer-related deaths and 18.1 million new cases worldwide in 2018 [1]. Breast cancer and

B. H. Woldegiorgis et al. (Eds.): ICAST 2022, LNICST 455, pp. 79–94, 2023.
https://doi.org/10.1007/978-3-031-28725-1_6

cervical cancer, which are the second and fourth most common cancer types in women globally, are the most common and lethal of all known cancer kinds [1].

According to statistics provided by the Addis Ababa City Cancer Registry Quadrennial Report of 2012–2015 investigation, women made up two-thirds of the city's cancer burden, with breast cancer accounting for the majority of that burden. Cervical cancer was the next most common cancer to affect the city's female population [2].

The two most prevalent cancers, breast cancer and cervical cancer, are also the two that kill the most women from cancer worldwide. Through improvements in early diagnosis methods and prescreening procedures, illness survival can be increased. A clinical examination, imaging-based screening, and pathological evaluation (a biopsy test) are often used diagnostic techniques.

Unfortunately, most underdeveloped nations [3–5] including Ethiopia have a physician and pathologist scarcity, which makes it difficult for many women with breast cancer and cervical cancer to receive an early diagnosis. Due to this, many women, especially those living in rural regions, miss out on the opportunity to receive an early diagnosis of breast cancer and cervical cancer [6]. Because of these illnesses, the death rate of mothers has increased. In addition, pathologists must review a lot of biopsy samples each day, which exhausts them. Moreover, the complexity of the cancer cells and their subjective decision, which is dependent on expert's performance, may lead to misdiagnosis of the subtypes and grade of the diseases from biopsy tests, which is essential to understand the biological characteristics and clinical behavior of the cancer cells.

Even though, Pap smear test, Biopsy test and colposcopy image tests are the common screening and diagnosing techniques for cervical cancer and breast cancer [7–9] they are sometimes prone to misdiagnosis. That is, their result can be either over interpreted or under interpreted. Over interpreted means that women without cancer are exposed to potentially harmful treatments and unnecessary expenses. On the other hand, inaccurate interpretation (under interpretation) of biopsy result could prevent women from getting the treatment early causing the cancer to grow more to invasive stage.

Now a days, machine learning and deep learning techniques are applied in different medical image and signal analysis works [10–18] in order to automate diagnosis systems and help physicians in getting support for making accurate decision. Artificial intelligence's (AI) application in the analysis of medical images can help increase diagnosis accuracy and reduce subjective variability and misdiagnosis rate by reducing the work load of pathologists.

In this project, an integrated breast cancer and cervical cancer diagnosis decision support system is developed in the form of web application.

2 Methodology

2.1 System Development and Technologies Used

When properly developed, clinical decision support systems can greatly improve the quality of treatment by delivering more accurate diagnoses, fewer mistakes, reduced costs, and more patient and provider satisfaction [19].

An automatic breast cancer and cervical cancer diagnosing system is developed and tested. The essence of this project is to develop an integrated web application which can be installed in any desktop or personal computer of users. It is designed as a decision support system for physicians. Furthermore, the system has many added features which can help researchers to acquire important information like statistical data of the prevalence of the diseases and organized labeled digital images for future studies can be acquired from the systems database.

The developed offline web application has the ability of classifying digital histopathological image of breast cancer in to binary (benign and malignant) classes and has the ability of classifying the given image into further eight subtypes (adenosis tumor, fibroadenoma tumor, phyllodus tumor and tubularadenoma tumor subtypes under the benign class and Ductal carcinoma, Lobular carcinoma, Mucinious Carcinoma and Papilary Carcinoma as a subtypes under the malignant cancer types), and classifies digital histopathological image of cervical cancer in to precqancerous cells, squamous carcinoma and adenocarcinoma. Furthermore, the developed system is capable of classifying pap smear images as normal and abnormal cells. Furthermore, using a colposcopy image as input, physicians can use the developed system to classify cervix type as type 1, type 2 and type 3. This information will help physicians to exactly know the transformation zone of the cervix.

To develop the overall system Django frame work was used. Django is an open source high-level python web framework that enables rapid development of secure and maintainable websites [20]. For front end development HTML, CSS, JavaScript, Bootstrap, and AJAX were used. For backend development python is used for storage SQLite database is used. Furthermore, the classification models were developed by using ResNet152 pretrained model. Finally, the Incremental Development Model was utilized to create the web application. This is a software development process in which a web application is developed, implemented, and tested gradually, with little adjustments made until the application is complete.

2.2 Image Classification System

2.2.1 Data

Microscopic histopathology images of breast cancer were collected from datasets provided by Jimma University Medical Center (JUMC), "break-his," [10] and "zendo" [11] repositories for the purposes of model training, validation, and testing. The images from JUMC were collected using an Opti-ka-vision camera attached to a simple light-microscope with four magnification powers (40X, 200X, and 400X), and a resolution of 2592 X 1936. The images were stained using the H&E staining process. Similar to how images from the break-his dataset were collected, different magnification factors were used to capture the images (40X, 100X, 200X, and 400X). The image frames were obtained from areas that had been affected by tumor growth.

A total of 915 Hematoxylin and Eosin (H&E) stained histopathology images from Jimma University Medical Center (JUMC) and St. Paul Hospital were obtained for the classification of cervical cancer. The OPTIKA light microscope with mounted digital camera, a smartphone camera, and a digital scanner was used to gather the data. 4x, 10x, 40x, and 100x magnification powers were used to capture microscopic images, with resolutions ranging from 419×407 to 2048×1536.

For the classification of cervix types, a total of 4005 pictures were gathered. After applying 5% acetic acid to the cervix, 133 colposcope images from Tercha General Hospital were obtained via a speculum (to reach the cervix) and 13MP and 18MP Tecno smartphone cameras. Three experienced gynecologists, including a senior medical doctor and an integrated emergency surgical officer (IESO), categorized the images after they had been gathered. The remaining pictures are from the Kaggle dataset, which is accessible to everyone [20]. Prior to data collection, Jimma Institute of Health's intuitional research review board granted ethical permission.

For testing and training, the Herlev Pap smear dataset (general public) was used. It includes 917 pictures of single cervical cells that have been classified and segmented using ground truth [16].

2.2.2 Classification Model

An image is recognized based on its visual information by the intricate image classification system used in computer vision. The accuracy of classification is primarily influenced by the characteristics of the dataset, the complexity of the analytical problem, and the efficiency of the classification method. Inference time, memory usage, computational complexity, recognition accuracy, and model complexity can all be used to assess how effective a classification method is. It is difficult to train a highly deep neural network because of the vanishing gradient problem. The gradient becomes exceedingly small as a result of repeated multiplication since it propagates back to earlier layers. Whenever the result of a network grows deeper, its performance also reaches a saturation point or soon drops [21, 22].

For the cancer classification tasks, the state-of-the-art pre-trained classification model, ResNet152, was used to classify all the multi class histopathology images, colposcopy images and Pap smear images. The model was trained with raw images acquired from the data repositories. Adam optimizer with a learning rate of 0.0001, maximum of 300 number of epochs, and loss function of sparse categorical cross entropy, were used for training. The models were validated using train-test split technique. Finally, new data were used to test the performance of the model for the different classification tasks.

2.3 Key Functions of the System

The web application provides several key functions including:

- **Doctor's login:** the doctor/ physician will be able to login after inputting the assigned username and password on the provided space.
- **Diagnosis:** the system is able to diagnose breast cancer and cervical cancer from histopathological images and screen cervical cancer from digital Pap smear images and colposcopy image.
- **Store patient information and data:** the system has the ability of saving all recorded and diagnosed cases for future use.
- **Statistics:** the system statistically provides the prevalence of the disease in the form of charts, so that researchers and health professionals can take necessary measures.
- **View and print results:** the doctor will be able to view all the diagnosed results and images and print out the results in case needed.

3 Result

3.1 Classification

For the BC binary classification, classification in to benign and malignant breast cancer types, better validation was achieved at the 120th epoch with a validation loss of 0.1284. at this epoch, 96.53% training accuracy and 94.62% validation accuracy. For the BC benign subtype classification, the lowest validation loss (0.4628) was achieved at the 170th epoch having a training accuracy of 82.14% and validation accuracy of 82.83%. For BC malignant subtype classification task, the lowest validation loss (0.1725) was achieved at the 276th epoch with a training accuracy of 97.7% and validation accuracy of 95.42%. Figure 1 indicates the model's accuracy and loss result obtained for BC

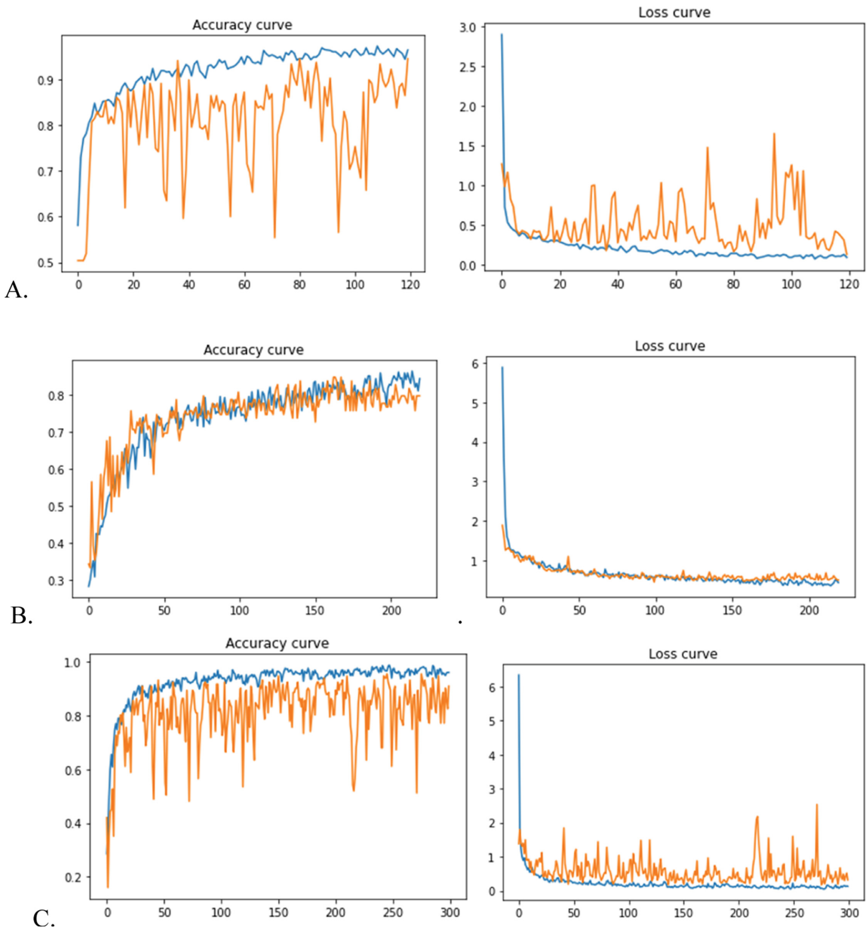

Fig. 1. Model accuracy (left) and loss plots (right) **A**. BC binary classification, **B.** Benign sub-type classification, **C.** BC Malignant subtype classification

histopathological image binary classification, BC benign subtype classification and BC malignant subtype classification.

Furthermore, the models were further evaluated using unseen test dataset. The test results for the trained models on the binary classification, benign subtype classification and malignant subtype classification are demonstrated below using the ROC plot in Fig. 2.

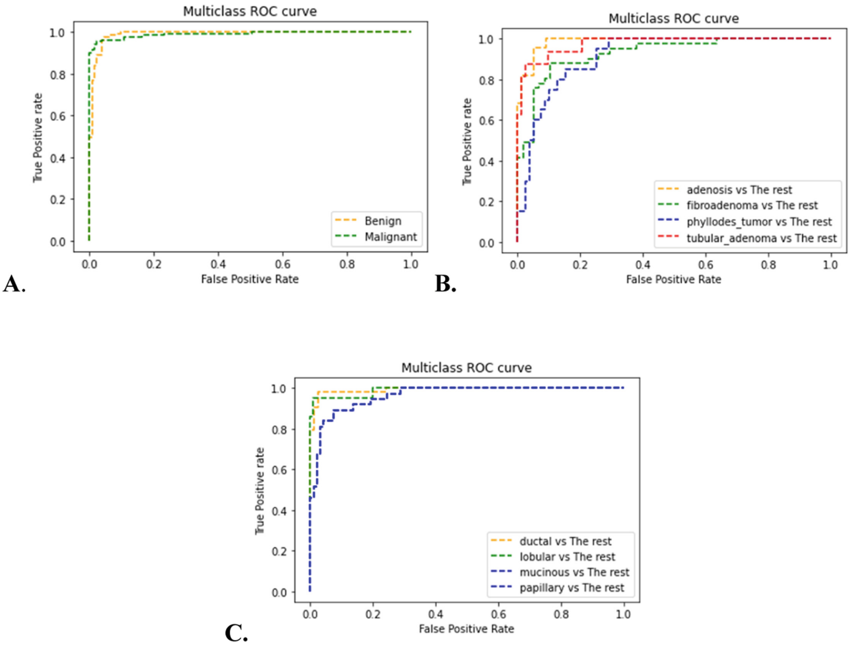

Fig. 2. ROC for BC **A.** Binary classification model, **B**. Binary subtype classification, **C.** Malignant subtype classification

For cervical cancer classification from histopathological image using ResNet152 the lowest validation loss (0.1515) was achieved at the 151th epoch with training accuracy 95.49% and validation accuracy of 96.17%. For cervix type classification the lowest validation loss of 0.7179 with training accuracy of 76.53% and validation accuracy of 67.91%. For pap smear classification the model achieved the lowest validation loss at the 173th epoch with a training accuracy of 86.1% and validation accuracy of 87.27%. Figure 3 shows training accuracy and loss results of ResNet152 model for classifying cervical cancer from histopathological images, cervix type classification task and pap smear image classification task respectively.

The trained models to perform cervix type classification and Pap smear digital image classification task were further evaluated using unseen test dataset. The test results for the models to perform the necessary classification are shown below using the ROC plot in Fig. 4.

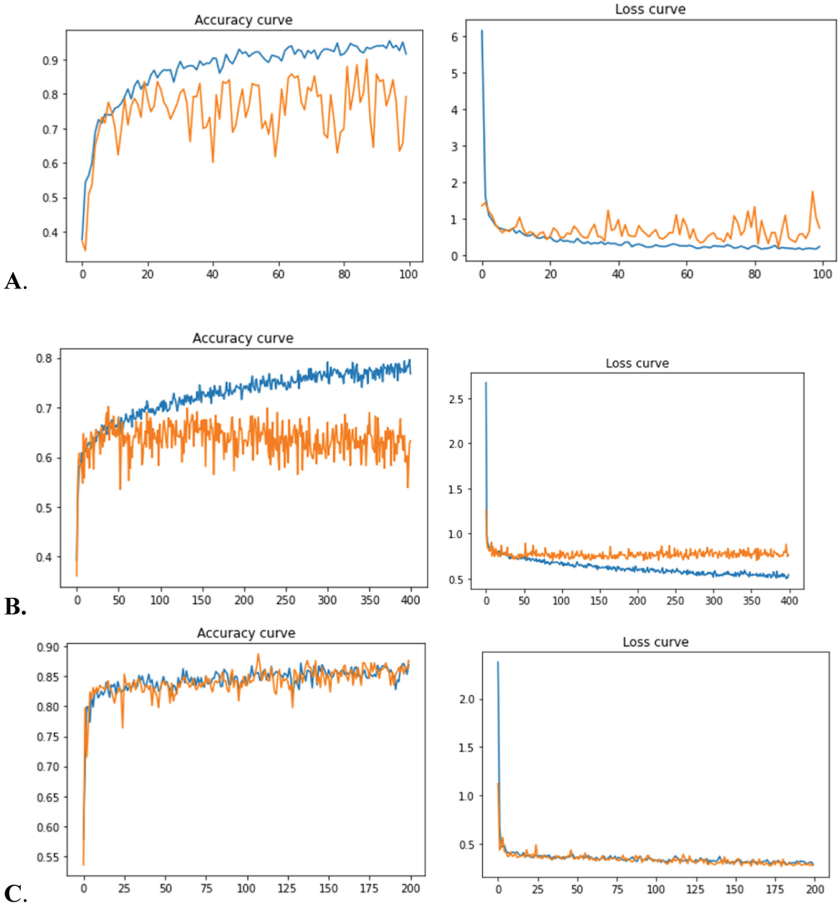

Fig. 3. Model accuracy (left) and loss plots (right) **A**. Cervical cancer histopathological image classification, **B.** Cervix type classification, **C.** Pap smear image classification

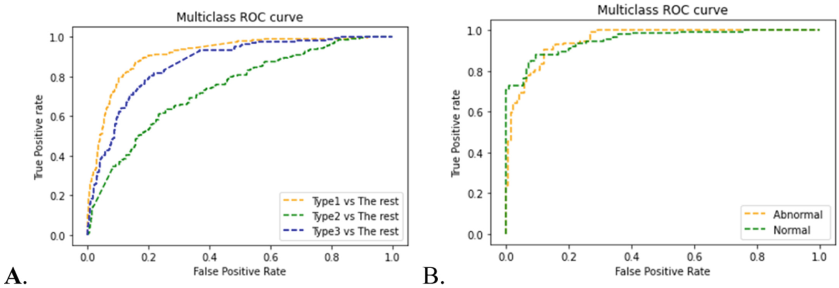

Fig. 4. ROC graph **A.** cervix type classification model, **B**. Pap smear image classification

3.2 Developed Web Page Result

The first interaction between the user and the web app is through the login page as indicated on Fig. 5. A registered doctor/physician who have been given a username and password by the admin of the system can login in to the system by inserting his/her unique ID /username and password. This page will help to restrict other people who are not given access to use the system from accessing patient's information on the web app. This will increase patient data confidentiality.

After a successful login by the doctor, the **homepage** will appear as shown in Fig. 6. This section provides brief information regarding Breast cancer and cervical cancer. This page also directs to different functions of the system like diagnosis, browsing patient records and images, statistical information etc.

As indicated in Fig. 7, **Diagnosis/screen** option on the left side of the home page will allow doctors to fill personal information about the patient and create new ID, if the patient is registering as a new patient. Or the doctor can use the patients ID to retrieve previously saved information and results of the patient. Then, the doctor can continue by selecting the type of analysis he/she is intending to do as demonstrated in Fig. 8. For example: if the doctor's intention is to do breast cancer diagnosis. Then he will choose **breast cancer** and click on "**select image for analysis**" this will direct the doctor to either capture new image or browse saved image from the computer. Figure 9 indicates the page where the doctor can upload or acquire image. Once the image is captured or browsed as shown in Fig. 10, the image will be displayed as indicated in Fig. 11. After that, by clicking the "**Start processing**", the system will start to analyze the class and subclass of the given image based on the deep learning model's knowledge. Finally, the result will be displayed in the form of percentage as indicated in Fig. 12. As the system is developed to be a decision support system for doctors, he/she will take the result of

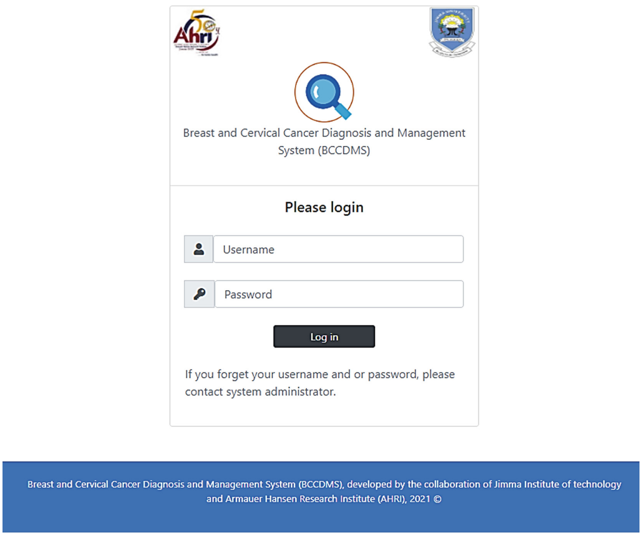

Fig. 5. Login page of BCCDMS

the system as an input and the final decision on the diagnosis result will be made by the doctor. Finally, the result achieved with the detail patient's information can be saved and printed as indicated in Fig. 13.

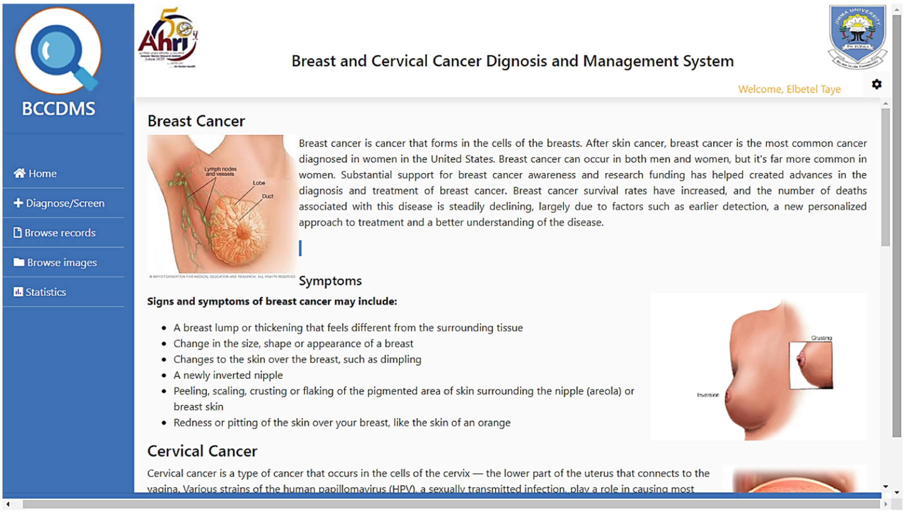

Fig. 6. Homepage of BCCDMS

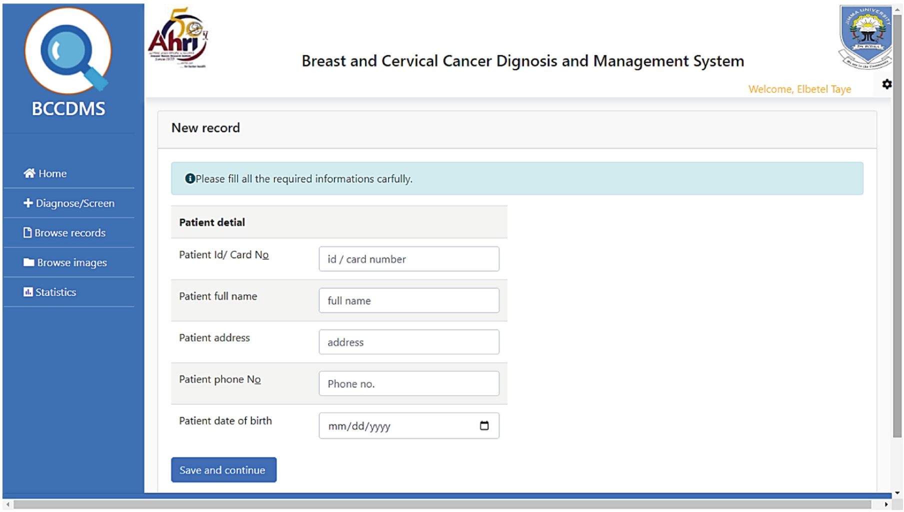

Fig. 7. Patient registration page

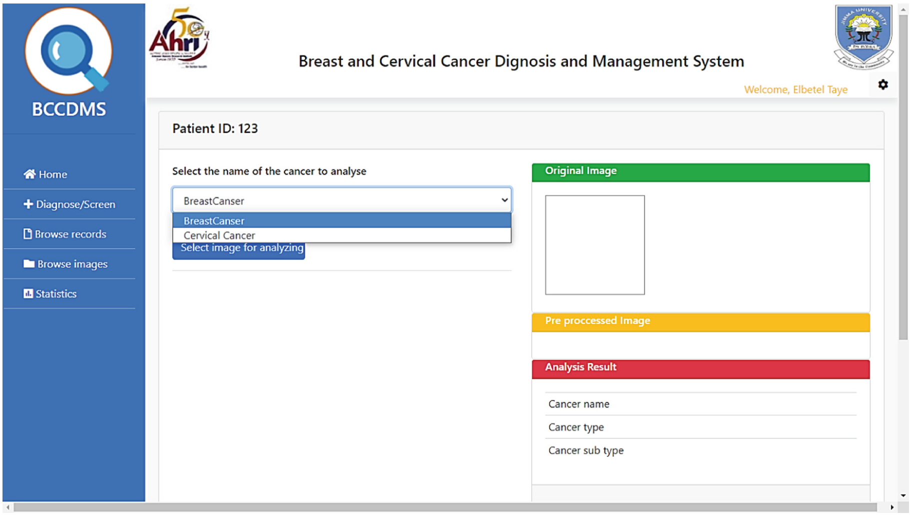

Fig. 8. Selecting the type of analysis

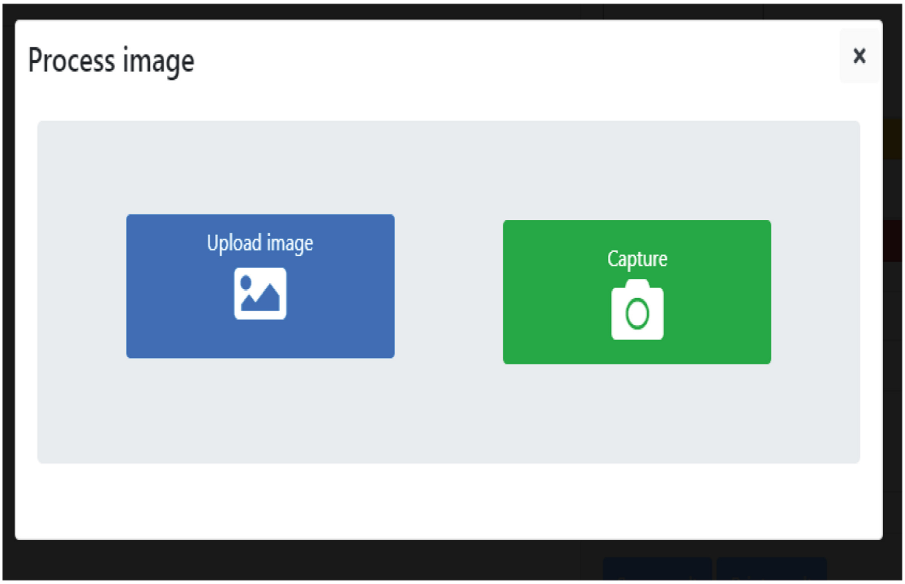

Fig. 9. Image Upload or Capture option

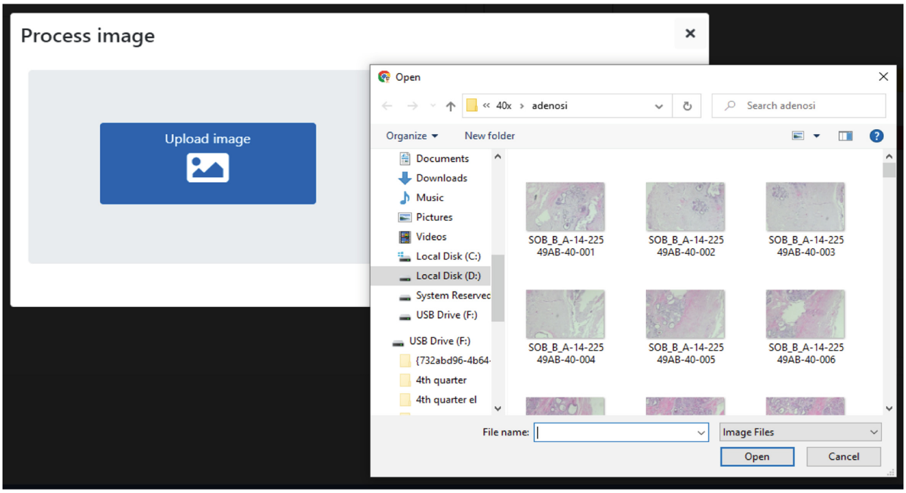

Fig. 10. Image Browsing

Fig. 11. Selected image processing

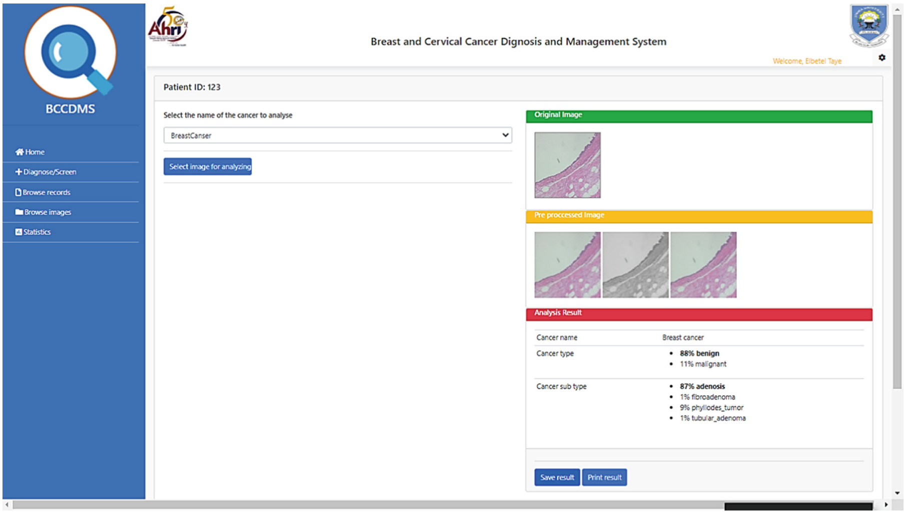

Fig. 12. Breast cancer analysis result

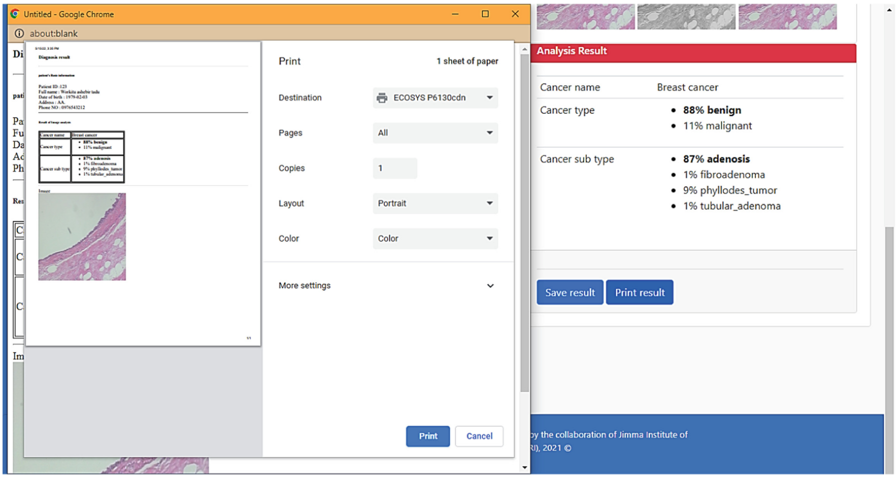

Fig. 13. Patient details and result printing Option

Furthermore, the system allows data retrieving through the **Browse data** option the available data can be filtered through the diagnosis type or just by inserting individual patients ID or name. This feature of the developed system will help researchers to find an organized digital dataset. Figure 14 shows each analyzed image saved under its labeled class/folder. Moreover, the system will help researchers and other concerned bodies to know how prevalent the diseases are in a given area. Figure 12 shows example of the chart indicating the frequency of the diseases from all up to date analyzed images (Fig. 15).

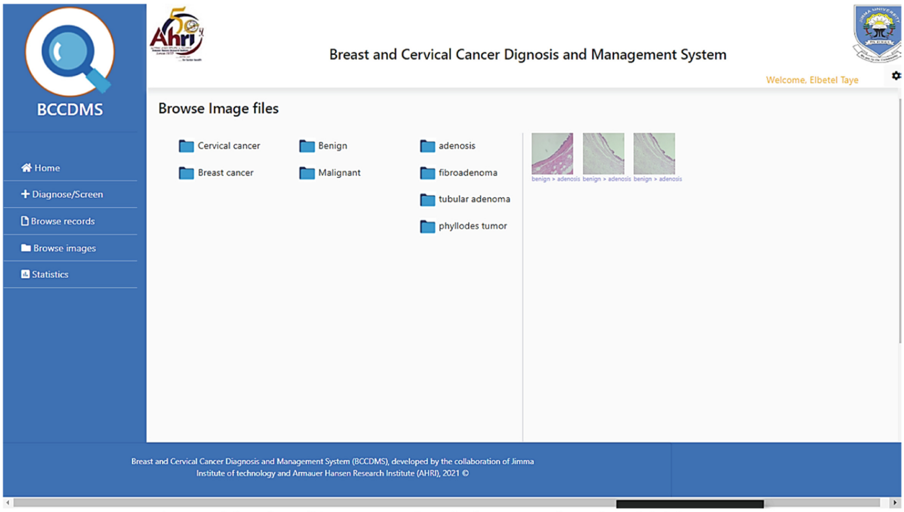

Fig. 14. List of images under their respective class/ folder

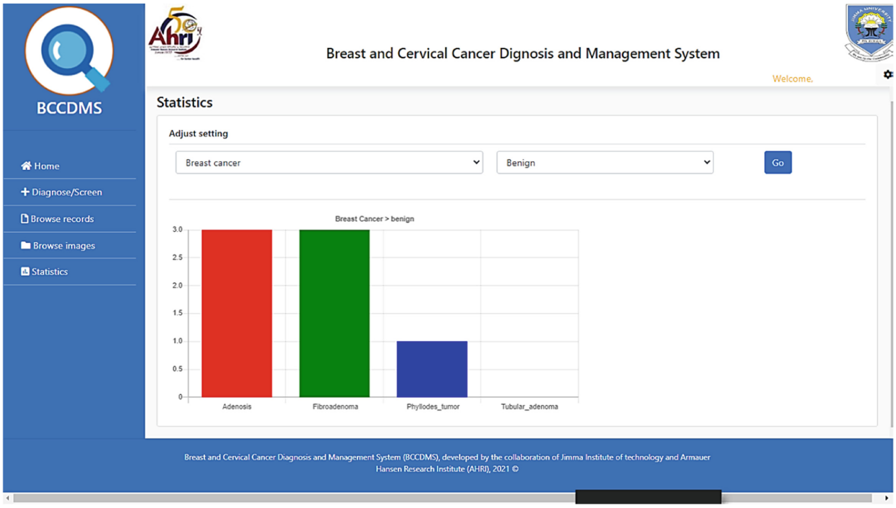

Fig. 15. Statistical representation of analyzed data

4 Discussion

In this paper, a web page was developed based on deep learning model in order to help pathologists and physicians in their decision making process when diagnosing breast cancer and cervical cancer.

For deep model training, validation and testing, histopathological images of breast cancer were obtained from BreakHis online dataset and locally acquired from Jimma

University Medical Center (JUMC). For the cervical cancer classification, histopathological images were acquired from JUMC, and St. Paul Hospital. For cervix type classification, 4,005 cervix colposcopy images were acquired from Tercha General Hospital. For the Pap smear image classification, a total of 1,417 images were collected from local health facilities and online public datasets. All images were pre-processed by applying data augmentation, image resizing and normalization techniques prior to model training.

For breast cancer binary classification (benign and malignant), ResNet152 model was trained for 120 epochs with 0.0001 learning rate, an Adam optimizer, and sparse_categorical_crossentropy loss function. A validation accuracy of 94.62% was achieved. Similarly for the benign subtype classification (Adenosis, Fibroadenoma, phyllodus, and tubularadenoma) the model was trained for 220 epochs and a validation accuracy of 82.83% was achieved. For the malignant breast cancer classification task (Ductal, Lobular, Mucinous, and papillary) the model was trained for 300 epochs and 95.42% validation accuracy was achieved. Finally, the trained models were tested and the satisfactory result was achieved as shown in the ROC in Fig. 2.

For the cervix type classification (Type 1, Type 2 and Type 3 cervix), the model (ResNet152) was trained for 400 epochs with similar learning rate, optimizer and loss function. A result of 76.53% validation accuracy was achieved. For the cervical cancer classification ('Squamous cell carcinoma', 'Precancer', 'Normal','Adinocarcinoma') from histopathological image using ResNet152 having the same parameters a validation accuracy of 96.17% was achieved. For Pap smear classification the model was trained for 200 epochs and a validation accuracy of 87.27% was attained. Finally, the models for the cancer classification, cervix type classification and Pap smear classification were tested with new images and a satisfactory result was achieved as shown in the ROC in Fig. 4.

All the classification models were integrated in one web based application for pathologists and physicians to perform the necessary tasks in an easy way. The developed web app as indicated in Fig. 5 has a login page that will restrict other people who are not given access to use the system from accessing patient's information on the web app. This will help to increase patient data confidentiality. After a successful login by the doctor, the homepage will appear as shown in Fig. 6. It has brief information regarding Breast cancer and cervical cancer. The page will also direct to different functions of the system like diagnosis, browsing patient records and images, statistical information etc.

This work presents an integrated diagnosis and screening system which has an advantage over the previous literatures [10–18] proposed in a way that it integrates both breast cancer and cervical cancer diagnosis and screening options to help in reducing misdiagnosis rate, fatigue on doctors and to increase the reliability of test results on the diagnosis. Besides, the developed system, through its option to save the images, has the ability to store the analyzed images with their result (the system will save the images as labeled images). This is one of the major contributions offered by the system to the existing health care system and the research world. As a result, anyone who wants to retrieve previous data can easily have access to it as indicated in Fig. 14.

In summary, in this paper, a full-fledged, integrated, magnification power independent and robust system is proposed for cervical cancer screening and diagnosis system by automating both the pre-screening (cervix type classification) and cervical cancer type classification tasks. We acknowledge that, staging of cervical cancer has not been included in this work. Moreover, adding more data may improve the accuracies of the proposed systems.

5 Conclusion

The proposed intelligent web-based tool is a robust decision support system designed to diagnose and manage breast and cervical cancer using multimodal image data including histopathological images, pap smear digital images and colposcopy images by employing the state-of-the-art artificial intelligence technique. The developed system will help in the reduction of cancer misdiagnosis, minimize the work load of physicians and improve management of breast and cervical cancer diagnosis. Moreover, the system has added a unique feature of storing labeled images to solve the existing problem of lack of organized labeled digital medical data. This will allow researchers and other health care professionals to easily retrieve digital images for further analysis and research.

References

1. World Health Organization Releases Latest Global Cancer Data - Cancer Health. https://www.cancerhealth.com/article/world-health-organization-releases-latest-global-cancer-data. Accessed 15 Dec 2019
2. Timotewos, G., et al.: First data from a population based cancer registry in Ethiopia. Cancer Epidemiol. **53**, 93–98 (2018)
3. Naicker, S., Plange-Rhule, .J, Tutt, R.C., Eastwood, J.B.: Shortage of healthcare workers in developing countries Africa. Ethn. Dis. **19**(1 Suppl 1), S1-60-4 (2009 Spring). PMID: 19484878
4. Legason, I.D., et al.: A protocol to clinically evaluate liquid biopsies as a tool to speed up diagnosis of children and young adults with aggressive infection-related lymphoma in East Africa "(AI-REAL)." BMC Cancer **22**(1), 1–9 (2022)
5. Kinfu, Y., Dal Poz, M.R., Mercer, H., Evans, D.B.: The health worker shortage in Africa: are enough physicians and nurses being trained? (2009)
6. Kisiangani, J., et al.: Determinants of breast cancer early detection for cues to expanded control and care: the lived experiences among women from Western Kenya. BMC Womens Health. **18**(1), 81 (2018). https://doi.org/10.1186/s12905-018-0571-7
7. Sornapudi, S., et al.: Automated cervical digitized histology whole-slide image analysis toolbox. J. Pathol. Inform. **12**(1), 26 (2021)
8. Safaeian, M., Solomon, D., Castle, P.E.: Cervical cancer prevention—cervical screening: science in evolution. Obstet. Gynecol. Clin. North Am. **34**(4), 739–760 (2007). https://doi.org/10.1016/j.ogc.2007.09.004
9. Veta, M., Pluim, J.P., Van Diest, P.J., Viergever, M.A.: Breast cancer histopathology image analysis: a review. IEEE Trans. Biomed. Eng. **61**(5), 1400–1411 (2014)
10. Spanhol, F.A., et al.: A dataset for breast cancer histopathological image classification. IEEE Trans. Biomed. Eng. **63**(7), 1455–1462 (2015)
11. Dimitropoulos, K., et al.: Grading of invasive breast carcinoma through Grassmannian VLAD encoding. PLoS ONE **12**(9), e0185110 (2017)
12. He, K., et al.: Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)

13. Alyafeai, Z., Ghouti, L.: A fully-automated deep learning pipeline for cervical cancer classification. Exper. Syst. Appl. **141**, 112951 (2020). https://doi.org/10.1016/j.eswa.2019.112951
14. Guo, P., et al.: Nuclei-based features for uterine cervical cancer histology image analysis with fusion- based classification. IEEE J Biomed Heal. Inf. **20**(6), 1595–1607 (2015). https://doi.org/10.1109/JBHI.2015.2483318
15. Zewde, E.T., Simegn, G.L.: Automatic diagnosis of breast cancer from histopathological images using deep learning technique. In: Berihun, M.L. (ed.) ICAST 2021. LNICSSITE, vol. 411, pp. 619–634. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-93709-6_42
16. Almubarak, H.A., et al.: A hybrid deep learning and handcrafted feature approach for cervical cancer digital histology image classification. Int. J. Healthc. Inf. Syst. Inform. **14**(2), 66–87 (2019). https://doi.org/10.4018/IJHISI.2019040105
17. Tian, Y., et al.: Computer-aided detection of squamous carcinoma of the cervix in whole slide images. Comput. Vis. Pattern Recognit. https://arxiv.org/abs/1905.10959
18. Wei, L., Gan, Q., Ji, T.: Cervical cancer histology image identification method based on texture and lesion area features. Comput. Assist. Surg. **22**(sup1), 186–199 (2017). https://doi.org/10.1080/24699322.2017.1389397
19. Zewdie, E.T., Tessema, A.W., Simegn, G.L.: Classification of breast cancer types, sub-types and grade from histopathological images using deep learning technique. Heal. Technol. **11**(6), 1277–1290 (2021). https://doi.org/10.1007/s12553-021-00592-0
20. Django, https://developer.mozilla.org/en-US/docs/Learn/Server-side/Django/Introduction Last Accessed: 15–5–2022
21. Cervical cancer dataset. https://www.kaggle.com/code/prakharpipersania/cervical-cancer. Accessed 17 Feb 2022
22. Gao, Z., Wang, L., Zhou, L., Zhang, J.: HEp-2 cell image classification with deep convolutional neural networks. IEEE J. Biomed. Health Inform. **21**(2), 416 (2017)