AI Doc Helper

Ranveer Kothavale[*1] *and Archana P. Ekbote*[2]

Abstract

For numerous times, numerous people have failed due to undetected conditions. Early discovery of these conditions at the micro bracket stage can be useful for furnishing proper treatment of the cases at an early stage and could have saved a lot of lives. A lot of exploration is being done to describe these conditions at the foremost. Thus, a computer-backed or Artificial Intelligence approach for detecting conditions at the early stage is being proposed, which makes use of machine, literacy and deep literacy algorithms for detecting conditions. This system will describe all general conditions similar to different types of cancer, malaria, diabetic retinopathy, etc. AI-DocHelper is being proposed as there's no system available that detects all these general conditions.

Keywords : **Artificial Intelligence, cancer, detection**

I. INTRODUCTION

Healthcare is one of the biggest issues in countries as it directly affects inhabitants' quality of life [1]. Despite this, the healthcare industry is primarily fragmented, universally scattered, and disjointed. From a clinical standpoint, providing appropriate patient care necessitates access to appropriate patient data, which is available infrequently when and where it is needed [2]. Furthermore, the vast range of test-ordering for different objectives emphasizes that sufficient data is needed for relevant test sets [3]. Idea has been extended by claiming that the vast discrepancies in general requesting to practicing pathology are largely due to individual diversity in clinical practice and, as a result, are possibly changeable through more harmonious and informed decision-making for croakers [4] [5].As a result, medical data frequently correspond to a large number of unrelated variables, such as demographics, complaint history, drug, disinclinations, biomarkers, medical images, or inheritable labels, all of which provide a different perspective on the status of a case. Furthermore, statistical parcels among the planned sources differ inherently. When analyzing similar data, experimenters and interpreters are confronted with two issues: the curse of dimensionality (the point space grows exponentially with the number of confines and samples) and the diversity of point sources and statistical parcels [6]. These variables cause detentions and trips during the complaint investigation, and as a result, cases are unable to receive appropriate care [7].As a result, medical, computational, and statistical sectors are faced with the task of finding new ways to describe prognosis and opinion of illnesses as standard paradigms fail to handle all of this data [8]. This demand is closely linked to advancements in other fields, such as Big Data (BD), Data Mining (DM), and Artificial Intelligence (AI) [9].

Manuscript Received : May 27, 2022 ; Revised : June 15, 2022 ; Accepted : June 20, 2022. Date of Publication : August 5, 2022.

R. Kothavale[*1], *Student,* Vidyavardhini's College of Engineering and Technology, K.T. Marg, Vartak College Campus Vasai Road, Vasai-Virar - 401 202, Maharashtra. Email : ranveerkothavale09@gmail.com ; ORCID iD: https://orcid.org/0000-0002-6800-3676

A. P. Ekbote[2], *Assistant Professor,* Vidyavardhini's College of Engineering and Technology, K.T. Marg, Vartak College Campus Vasai Road, Vasai-Virar-401 202, Maharashtra. Email : archana.ekbote@vcet.edu.in ; ORCID iD: https://orcid.org/0000-0001-8342-3612

DOI : https://doi.org/10.17010/ijcs/2022/v7/i4/172377

II. PROBLEM STATEMENT

AI-DocHelper, an Artificial Intelligence grounded system that can determine general conditions at an early stage, will take textual and/ or image data as input. All applicable features will be uprooted from the input data. For a complaint to be detected, different bracket models can be applied to the point uprooted training data. The model that gives stylish precision and minimal loss in the testing phase will be named the final model for detecting a particular complaint. The affair or final result is the bracket of whether a particular case has a particular complaint or not.

III. LITERATURE SURVEY

A survey of the existing systems was done and the limitations were noted. Various researches have been conducted towards predicting results of the disease.

Malaria Detection Using Multiple Deep Learning Approaches

For all the image data or datasets available for illness discovery, our system will make use of Convolutional Neural Networks, which are a part of deep literacy [10]. All the image data will be preprocessed to prize all the applicable features for the discovery of a particular complaint. All the inapplicable features or data will be removed from the images. The inapplicable data will be removed because it will affect the delicacy of the model and will make it delicate for the presence of that particular complaint [8].

Breast Cancer Detection in Medical Images.

The following passageways make up a broad system framework for bone cancer discovery:

↳ **Positioning of the image:** Imaging artifacts and inconsistencies induced by various imaging settings may have a significant impact on the next step in the discovery process. For better discovery performance, it is necessary to remove the variability and vestiges with image proposing methods.

↳ **Area segmentation for the Region of Interest (ROI):** We must reward the most applicable corridor of the image before executing discovery Styles on it because we only observe the applicable sections of the entire slide image during discovery.

↳ **Extraction of features:** Due to the high constraints of raw picture data, it is impossible to use them directly for brackets. Point birth might convert raw image data into a point space with far less constraints, making it better suited to the bracket problem.

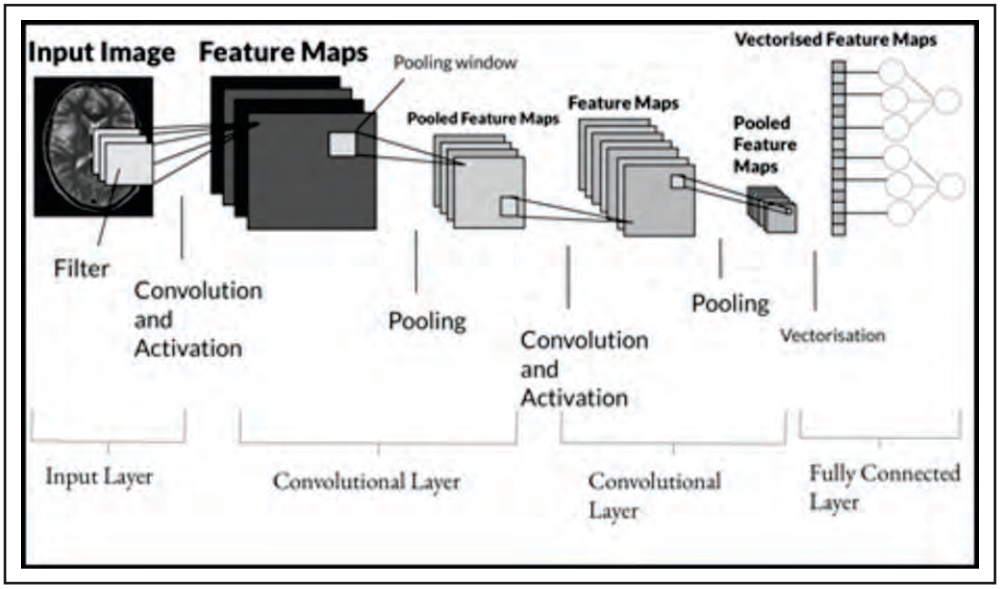

Fig .1. Malaria Detection Model

IV. PROPOSED SYSTEM

AI-DocHelper, an Artificial Intelligence predicated system that can describe general conditions at an early stage, will take textual and/ or image data as input. All applicable features will be uprooted from the input data. For a complaint to be detected, different bracket models will be applied to the point uprooted training data. The model that gives the stylish delicacy and minimal loss in the testing phase will be named as the final model for detecting a particular complaint. The affair or final result is the bracket of whether a particular case has a particular complaint or not.

V. IMPLEMENTATION

For conditions, conduct research for the textual and image dataset. All of the data cleaning and preprocessing has been completed. The systems were checked, and the limits were documented. On the image and textbook datasets, all of the algorithms will be tested, and the model with the most stylish delicacy and least loss will be named. All datasets will be trained on the model once the algorithm has been identified. The model will be preserved for future use as well. On the perfected design, a user-friendly device can be imposed. The system will be tested after the front end and reverse end have been connected. All of the system's features will be put to test. The model will be ready for deployment after bugs are fixed. The server will host the GUI as well as the database. Both the user interface and the backend will be tested. Bugs will be addressed. All of the bugs have been addressed, and testing will begin. The final phase of testing is now complete.

Brain Tumor Classification (MRI) Dataset

Tumor in brain is the most serious cause in children as well as various adults. About 85-90% of Brain tumor accounts in the Central Nervous System, i.e. CNS. Each year around 11,700 patients get diagnosed with Brain tumor. Men have only a 5 year survival rate of roughly

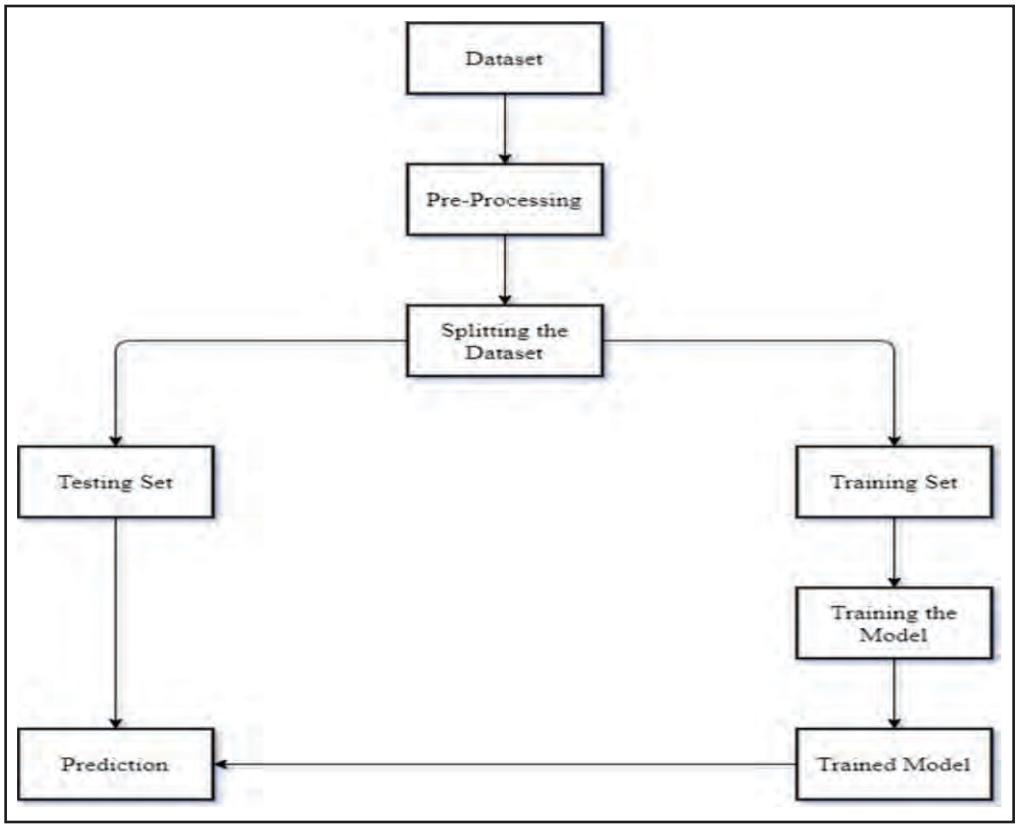

Fig. 2. Block Diagram of Training the System.

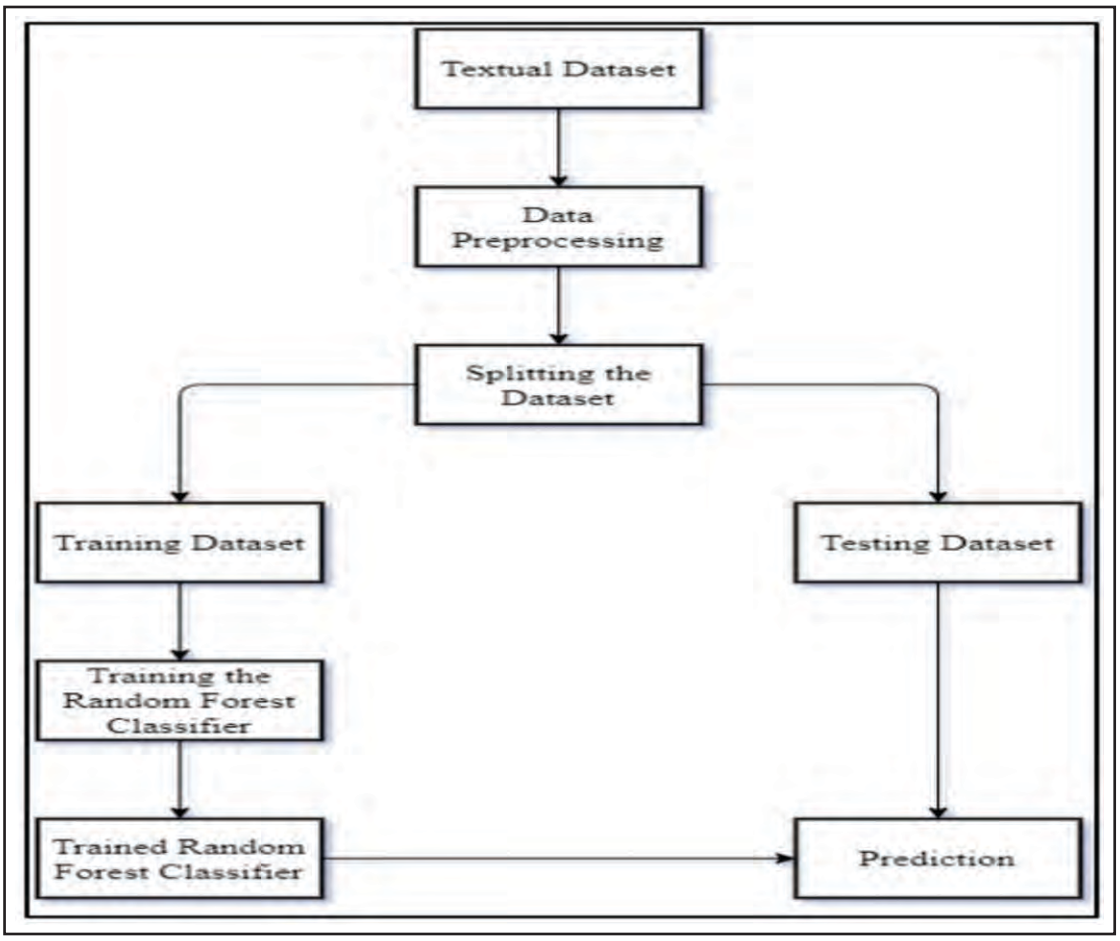

Fig. 3. Workflow for Textual Dataset

34% and women have survival rate of around 36 %when they have serious brain or CNS cancer [12]. Cerebrum Tumors are referred to as pituitary tumors and other names. Planning, correct treatment, and proper diagnostics are required to improve long-term chances of patients.The most effective approach for measuring brain growth is magnetic resonance imaging (MRI). The outputs produce an enormous amount of image data. These images are examined by a radiologist. Human assessment can be prone to mistakes due to the degree of complications associated with brain growths and their qualities.

No. of Samples : 3264

Classes : **No Tumor, Glioma Tumor, Meningioma Tumor, Pituitary Tumor**

Breast Cancer Dataset

Cancer growth takes place in the cells of womens' breasts. Breasts grow in ladies. Symptoms of Breast Cancer are irregularity in breast, minor releases from areola and frequent changes in shape and size [11]. The therapy depends on the development stage. This might comprise of radiation, chemotherapy, chemical treatment, and procedure in medical.

No. of Samples : 1578 samples

Classes : Normal, Benign, Malignant

Chronic Kidney Disease Dataset

Our kidneys perform an important function to help filter blood and pass waste as urine. Chronic order complaint, also referred to as chronic disorder describes the gradational loss of this function. At advanced stages, fluids having dangerous situations, waste and electrolytes can make up in the body. Once this happens, cases must go through dialysis or consider a transplant [13]. In this design we see if we can prognosticate if a case will have

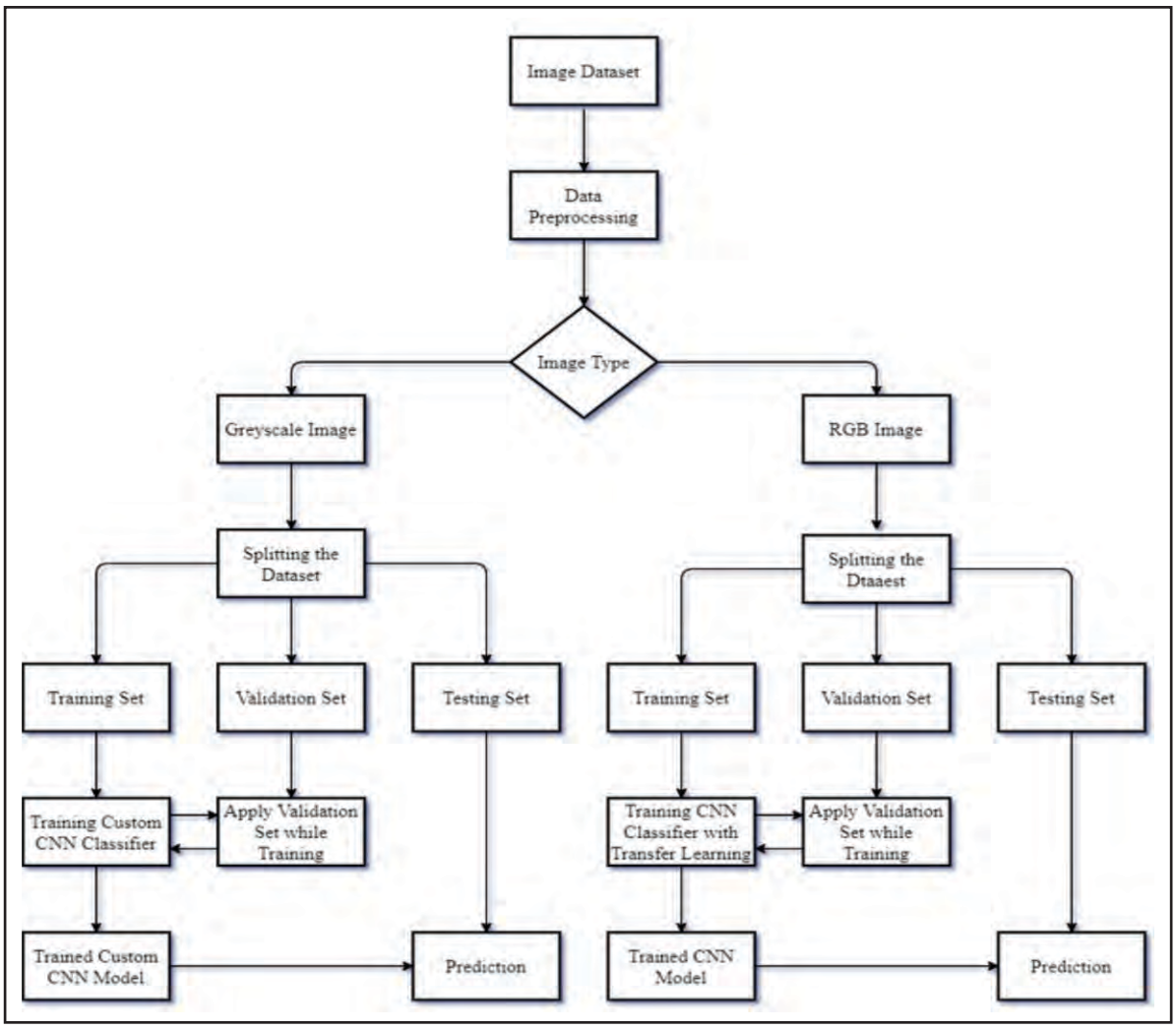

Fig. 4. Workflow for Image Dataset

chronic kidney complaints or not using 24 predictors. However, it may be suitable to describe and help cases at threat. If we're suitable to find variables with a strong influence on order failure [9].

Attribute Information :

- Age (in years)
- BP (in mm of Hg)
- Su (Sugar)
- RBC (Red Blood Cells)
- Al (Albumin)
- Sg(Specific Gravity)
- Ba(Bacteria)
- PC (Pus Cell)
- PCC(Pus Cell Clumps)

VI. RESULTS

This dataset was processed and fed to numerous machine literacy and deep literacy models similar to the K-Nearest Neighbour, Logistic Regression, Decision Tree , Naive Bayes, Random Forest, Support Vector Logistic Regression, K-Nearest Neighbor, Support Vector Classifier, Naïve Bayes, and Artificial Neural Network0 to decide which machine or deep literacy model gives the stylish delicacy and has the least loss.

TABLE I.
COMPARISON OF ACCURACY FOR TEXTUAL DATASET

MODEL	ACCURACY
Logistic Regression	82.46%
K-Nearest Neighbour Classifier	79.87%
Support Vector Classifier	82.46%
Naïve Bayes	79.22%
Decision Tree Classifier	70.77%
Random Forest Classifier	82.46%
Artificial Neural Network	70.77%

TABLE II.
ACCURACY AND LOSS OF ALL MODELS

Disease Name	Training Accuracy	Training Loss	Testing Accuracy	Testing Loss
Breast Cancer	87.36%	31.24%	85.49%	38.67%
Pneumonia	95.86%	10.92%	98.25%	14.88%
Glioma Tumor	92.20%	19.64%	92.41%	19.73%
Meningioma Tumor	93.19%	18.18%	90.52%	21.80%
Diabetes	0.8240	0.2310	0.8052	0.2513
Pituitary Tumor	99.06%	3.33%	98.85%	3.03%
Melanoma (Skin) Cancer	85.12%	31.37%	83.87%	33.26%
Acute Lymphoblastic Leukemia (ALL)	84.14%	37.54%	70.43%	58.85%
Heart Disease	0.9137	0.2513	0.8897	0.2883
Chronic Kidney	0.8963	0.1953	0.8663	0.2153
Malaria	89.69%	26.80%	90.49%	24.19%
Brain Tumor	89.21%	28.49%	66.67%	33.72%
Lung Cnacer	94.41%	13.80%	95.26%	11.60%
Invasive Ductal Carcinoma (IDC)	80.57%	43.16%	78.70%	47.90%
Tuberculosis	96.69%	9.75%	95.64%	11.64%
COVID-19	86.56%	33.12%	86.90%	32.88%

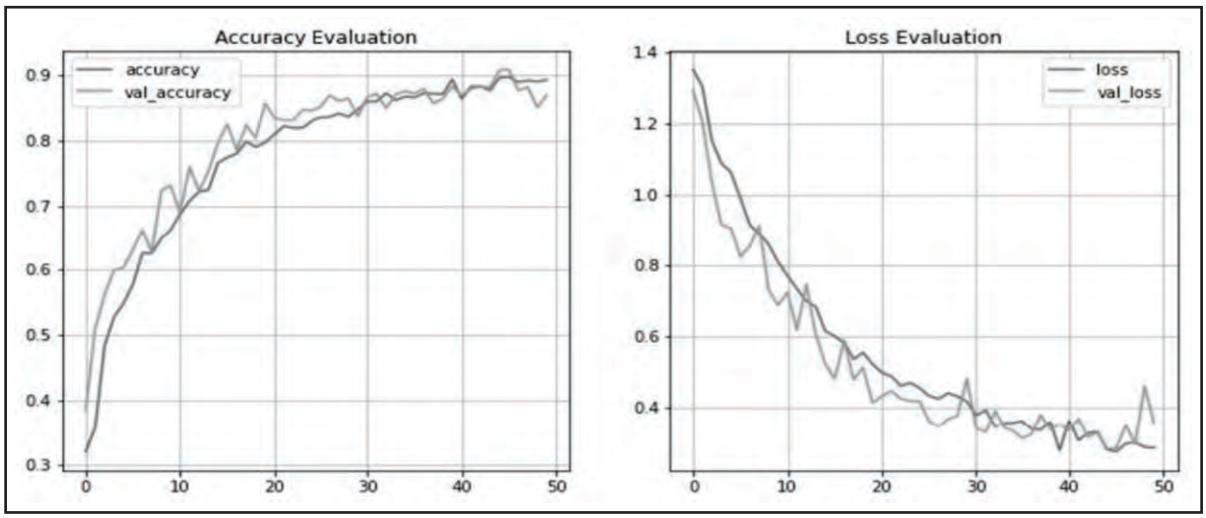

Fig. 5. Brain Tumor Classification

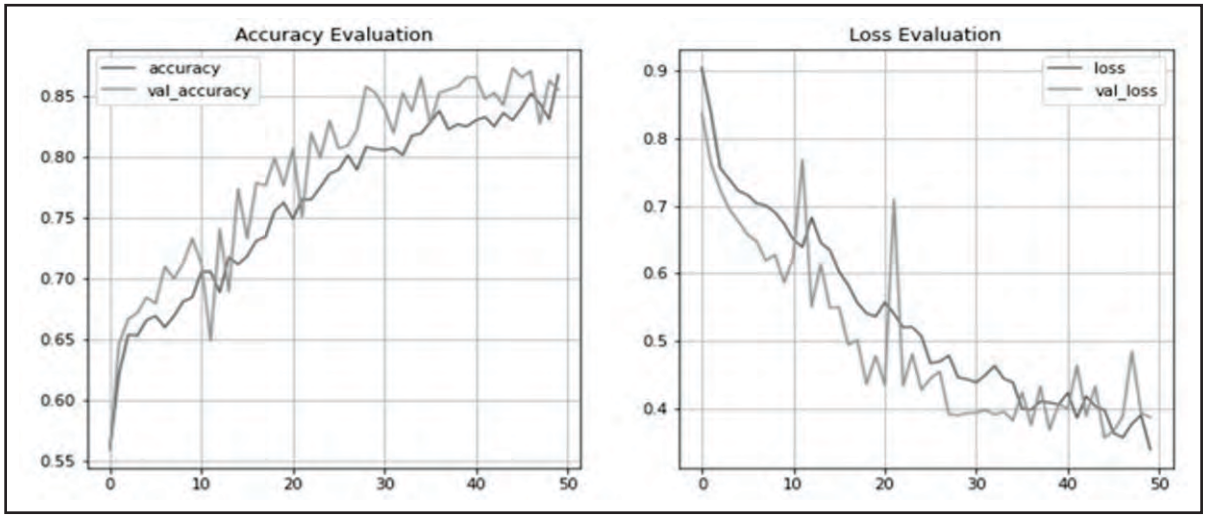

Fig. 6. Breast Cancer Detection

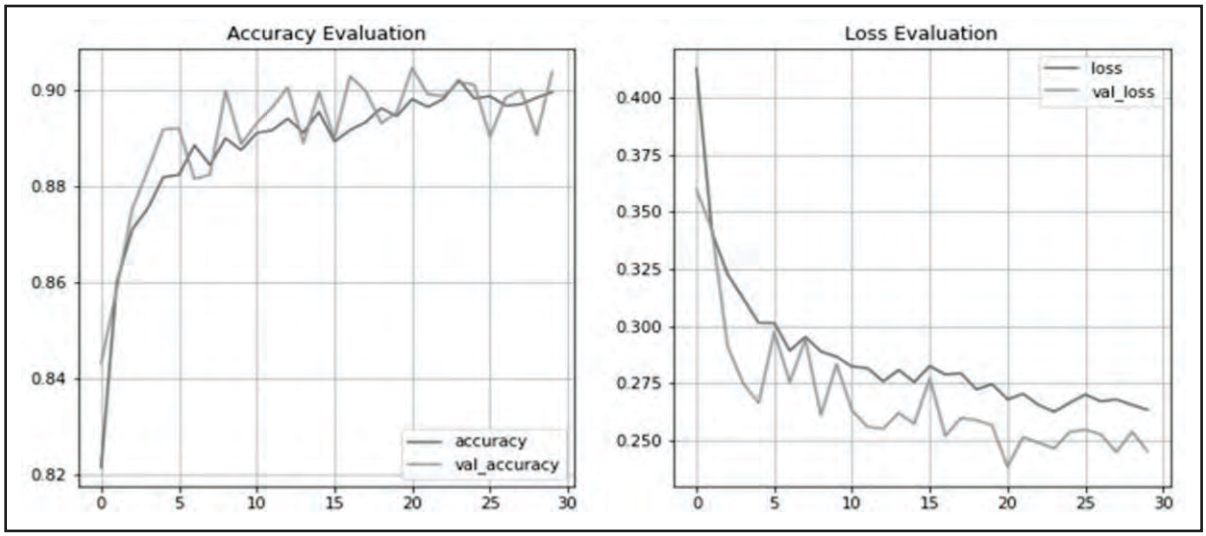

Fig. 7. Malaria Detection

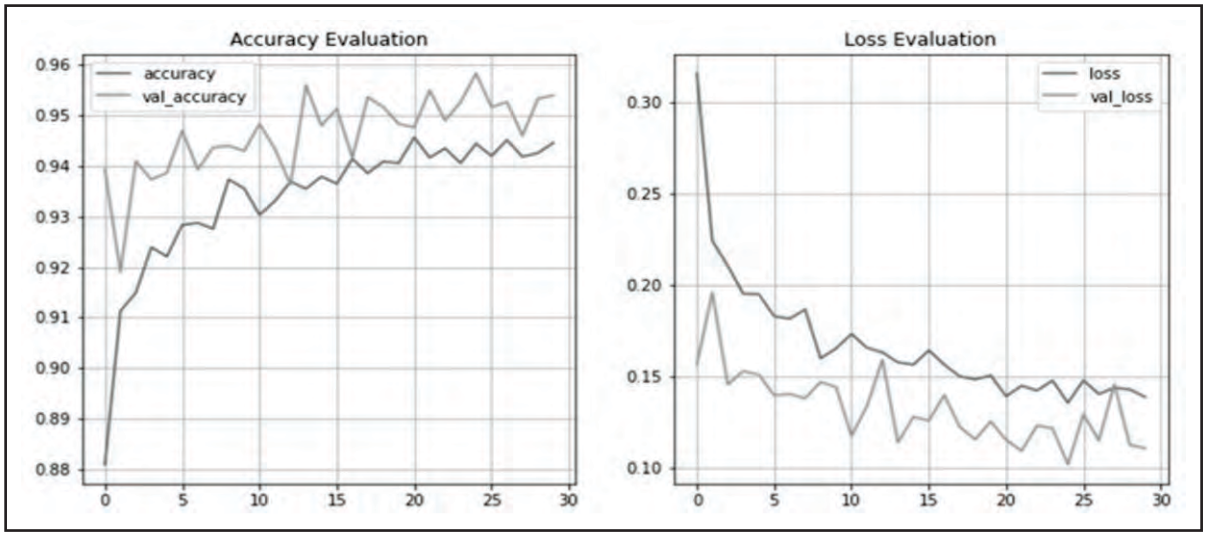

Fig. 8. Lung Cancer Detection

VII. CONCLUSION

The research was done, and no system was found available. The available systems had restrictions that they were suitable to find a particular type of disease. Available systems are confined to finding only one or two diseases. AI-DocHelper is better than numerous systems available for the request as it can describe numerous conditions. Once the system is available in the request, it will help to save numerous lives with perfection. Mortal errors will be reduced to a great extent. A patient can plan the course of treatment for a case veritably snappily, as the system gives the results incontinently, as compared to ultramodern systems available.

Once further datasets are available, they can be trained, and new conditions can be added to the system. The discovery models can be trained using further data, to increase their delicacy and the loss. The more the model has trained the more its properties for the discovery of issues will be fine-tuned.

AUTHORS' CONTRIBUTION

The dataset was cleaned and modified by Ranveer Kothavale. Ranveer Kothavale scraped data from the internet and added it to the dataset. Then he used various models and did the GUI work. Dr. Archana Ekbote provided timely guidance and assistance when required.

CONFLICT OF INTEREST

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in the manuscript.

FUNDING ACKNOWLEDGEMENT

The authors have not received any financial support for the research, authorship, and/or for the publication of the article.

REFERENCES

[1] P. Bagga and R. Hans, "Applications of mobile agents in healthcare domain: A literature survey," *Int. J. Grid Distribution Comput.,* vol. 8, no. 5, pp. 55–72, 2015. [Online]. Available: http://article.nadiapub.com/IJGDC/vol8_no5/5.pdf

[2] Z. Y. Zhuang, L. Churilov, F. Burstein, and K. Sikaris, "Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners," *Eur. J. Oper. Res.,* vol. 195, no. 3, pp. 662–675, 2009, doi: 10.1016/j.ejor.2007.11.003.

[3] R. S. Dick, E. B. Steen, and D. E. Detmer, *The computer-based patient record: An essential technology for health care.* National Academies Press-Washington DC, USA, 1997.

[4] J. E. Wennberg, "Dealing with medical practice variations: A proposal for action," *Health Affairs,* vol. 3, no. 2, pp. 6–32, 1984. https://pubmed.ncbi.nlm.nih.gov/6432667/

[5] W.S.A. Smellie, D. Wilson, C. A. M. McNulty, J. J., W. Irvine, P. C. Dore, G. Handley, M. J. Galloway, G. A.. Spickett, D. I. Finnigan, D. A. Bareford, M. A. Greig, and J. Richards, "Best practice in primary care pathology: Review 1," *J. Clin. Pathology,* vol. 58, no. 10, pp. 1016–1024, 2005, doi: 10.1136/jcp.2004.025049.

[6] M. Daniels and S. A. Schroeder, "Variation among physicians in the use of laboratory tests II. Relation to clinical productivity and outcomes of care," *Med. Care*, vol. 15, no. 6, , pp. 482–487, 1977, doi: 10.1097/00005650-197706000-00004.

[7] P. J. Stuart, S. Crooks, and M. Porton, "An interventional program for diagnostic testing in the emergency department," *Med. J.,* vol. 177, no.3, pp. 131–134, 2002, doi: 10.5694/j.1326-5377.2002.tb04697.x.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: arxiv.org/abs/1409.1556.

[9] R. Casanova, S. Saldana, E. Y. Chew, R. P. Danis, C. M. Greven, and W. T. Ambrosius, "Application of random forests methods to diabetic retinopathy classification analyses," *Plos One,* vol. 9, no. 6, p. e98587, 2014, doi: 10.1371/journal.pone.0098587.

[10] "Malaria cells image dataset," [Online]. Available: https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria

[11] https://scholar.cu.edu.eg/?q=fahmy/pages/dataset

[12] "Brain Tumor Classification (MRI)," [Online]. Available: https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri

[13] "Chronic Kidney Disease Dataset," [Online]. Available: [https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease

About the Authors

Ranveer Kothavale is pursuing B.E.(Information Technology). His research interests include Machine Learning and Data Analytics.

Archana P. Ekbote is Assistant Professor with the Department of Information Technology at Vidyavardhini's College of Engineering & Technology, University of Mumbai, India. Her research interests include signal and image processing, and pattern recognition.