iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50652

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Detection of Cancer Using Boosting Tech Web App

Mareedu Girish[1], Madasu Satish[2], Thota Hemanth[3]
[1, 2, 3]*Bharath Institute of Higher Education and Research*

Abstract: ***The early detection and prognosis of a cancer type have turned into a major requirement, as it facilitates successive medical treatment of patients. The machine learning field has shown greater potential in applications such as disease prediction and drug response prediction. Input is obtained in the form of an image for cancer prediction. Output results are acquired instantly in real time. We will be using CNN methodology. The existing systems are simple and effective but are extremely vulnerable to impact. Moreover, state-of-the-art methods work on just one algorithm which makes it less accurate and more time-consuming. We propose an end-end web application that predicts cancer using distinct techniques related to deep learning... The Advantages of the proposed system are that it could be the very first-of-its-kind, cost efficient, and highly accurate application that provides complete and accurate cancer prediction. The proposed application is highly applicable in the classification, and diagnosis of cancer and tumor diseases and is expected to become more important in medical practice shortly***
Keywords: Deep Learning, Artificial Intelligence, Cancer prediction,CNN

I. INTRODUCTION

Over the past decades, a continuous evolution related to cancer research has been performed. Scientists applied different methods, such as screening in the early stage, in order to find types of cancer before they cause symptoms. Moreover, they have developed new strategies for the early prediction of cancer treatment outcomes. With the advent of new technologies in the field of medicine, large amounts of cancer data have been collected and are available to the medical research community. However, the accurate prediction of a disease outcome is one of the most interesting and challenging tasks for physicians. As a result, ML methods have become a popular tool for medical researchers. These techniques can discover and identify patterns and relationships between them, from complex datasets, while they are able to effectively predict future outcomes of a cancer type. Given the significance of personalized medicine and the growing trend in the application of ML techniques, we here present a review of studies that make use of these methods regarding cancer prediction and prognosis. In these studies, prognostic and predictive features are considered which may be independent of a certain treatment or are integrated in order to guide therapy for cancer patients, respectively. In addition, we discuss the types of ML methods being used, the types of data they integrate, and the overall performance of each proposed scheme while we also discuss their pros and cons. An obvious trend in the proposed works includes the integration of mixed data, such as clinical and genomic. However, a common problem that we noticed in several works is the lack of external validation or testing regarding the predictive performance of their models. It is clear that the application of ML methods could improve the accuracy of cancer susceptibility, recurrence, and survival prediction. Based on this, the accuracy of cancer prediction outcomes has significantly improved by 15%–20% in the last years, with the application of ML techniques.

II. LITERATURE REVIEW

SVM, CNN, and KNN, three algorithms for predicting the outcome of breast cancer, were examined using various datasets in

1) All of the experiments are run using PyCharm and the Anaconda platform in a simulation environment. There are three types of research objectives. Cancer prediction is the first domain, followed by diagnostic and therapy forecast, and treatment outcome prediction is the third domain.
2) When GAN-enhanced feature learning is paired with hybrid training employing the ROI and the full picture, better classification performance and an effective end-to-end scheme are achieved.
3) The authors employed the (GAN) in this research to create synthetic mammographic images from the Digital Database for Screening Mammography (DDSM). Using the DDSM, we extracted two sets of regions of interest (ROIs) from the images: normal and anomalous (cancer/tumor). These ROIs were used to train the GAN, which then generated artificial images.
4) Traditional augmentation approaches have a lot of limitations, especially in situations where the images must meet strict criteria, such as medical datasets.
5) Traditional portable devices have numerous defects, such as comfort for long-term use and deficient quality, etc. Therefore, health observance is done with conventional portable devices are difficult to be sustainable.

©IJRASET: All Rights are Reserved

6) In the United States, electronic medical records are quickly growing, allowing for a significant increase in the amount of clinical data that may be obtained electronically. Probably, there has been rapid advancement in clinical analysis tools to analyse huge volumes of knowledge and derive new notions from that synthesis, which is a subset of what is known as big data.

7) A small number of risk factors for substantial falls show consistently, including disdain for the variety of settings, pace imbalance, stimulated confusion, bladder problems frequency, a history of falls, and administration of "guilty" medicines (particularly sedatives/hypnotics).

8) The effectiveness of supervised and unsupervised models for breast cancer categorization is examined in this research. This paper makes use of data from the Wisconsin Breast Cancer Dataset. Scaling and main component analysis are used to choose features. The Ensemble Voting technique is appropriate as a forecasting model for breast cancer, according to the final results. There are 569 instances of breast cancer in the raw data.

9) In this study, we compare the predictive performance, area under the receiver operating (AUC), and performance parameters of multiple machine learning algorithms for breast cancer prediction. The Wisconsin Dataset of Breast Cancer is being used for simulation purposes (WDBC).

10) In this paper, AI computations are used to predict the occurrence of breast cancer in women. The demonstration of AI calculations is graded on their expected correctness. On a regular informational collection, the four AI computations are applied. The Support vector Machine technique is found to be the best for breast cancer illness prediction after being run in the Python computer language.

III. EXISTING SYSTEM

The early detection of cancer has been reported to increase the survival rates and successful treatment. Although prediction results achieved are promising, the traditional models built on primitive approaches are still far from being highly accurate and efficient. Moreover, state-of-the-art methods are built on just one algorithm instead of using a multi-modal approach. Thus, the outputs predicted can be highly inaccurate while some severe conditions may go completely undetected. This could lead practitioners to false assumptions and improper diagnoses and treatments provided to patients.

IV. PROPOSED SYSTEM

We proposed a novel mechanism for detecting cancer from the given input image by applying deep learning algorithms especially. The aim of developing this application with the help of deep learning algorithms is to immensely help to solve health-related issues by assisting physicians to predict and diagnose cancer at an early stage. It also solves the problem of survivability prediction in clinical databases. It can analyze huge datasets and also find hidden or unexpected correlations among diverse attributes. The accurate analysis of our proposed application benefits early disease prediction, patient care, and community services.The overall accuracy of the proposed scheme has been evaluated with the traditional state-of-the-art models and the results from our proposed application show a higher accuracy rate.

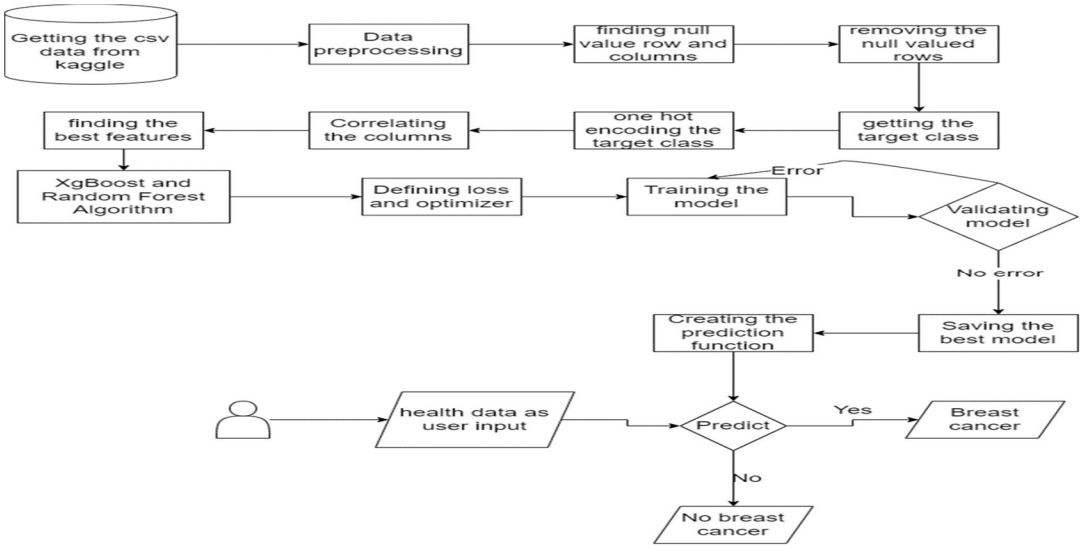


Fig 1: Architecture diagram

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

V. MODULES

1) Module 1: Data Acquisition and Preprocessing

Data preprocessing is a crucial step in the machine learning pipeline that involves transforming raw data into a usable format for machine learning models. This process is essential because real-world data is often incomplete, inconsistent, and noisy. Preprocessing data involves cleaning, transforming, and integrating data from multiple sources.

The first step in data preprocessing is data cleaning, which involves removing missing values, outliers, and irrelevant data. Missing values can be dealt with by either removing the row or filling in the missing value with an appropriate value like the mean, median or mode. Outliers can be detected and removed using statistical techniques such as Z-score analysis or using domain-specific knowledge. Irrelevant data can be removed by selecting only relevant features or using feature extraction techniques like principal component analysis (PCA).

The second step in data preprocessing is data transformation. This involves converting data into a standard format that can be used by machine learning algorithms. For instance, categorical variables need to be transformed into numerical variables through encoding techniques like one-hot encoding or ordinal encoding.

The third step in data preprocessing is data integration, where data from different sources are combined to create a unified dataset. This is achieved through data linking and data merging techniques.

The final step in data preprocessing is data reduction, which involves reducing the dimensionality of the dataset to remove redundant features that do not contribute to the predictive accuracy of the model. This can be achieved through feature selection techniques like recursive feature elimination (RFE) or dimensionality reduction techniques like principal component analysis (PCA).

2) Module 2: Implementation Of Model

In our research report, we implemented two machine learning algorithms: XGBoost and Random Forest. XGBoost uses a sequential process to generate decision trees, with unreliable predictors given more weight in the first decision tree to inform subsequent ones. Random Forest constructs a "forest" of decision trees, introducing more randomness into the model and producing unbiased results. We used binary cross entropy loss function to compare target and expected output values and minimize loss, and the Stochastic Gradient Descent optimizer to adjust hyperparameters and minimize error.

We split the dataset into training, validation, and test sets. The training set was used to train the model, the validation set was used to validate performance during training, and the test set was used to test the model after training.

The goal of splitting the dataset was to prevent overfitting and ensure that the model could accurately classify samples it had not seen before.

Finally, we saved the best model to reuse it in the future without having to form the model again, which could affect productivity. Saving the model while it is being trained facilitates model comparison to determine which champion model to use in production.

3) MODULE 3: Creating a Web App

An important part of building a machine learning model is to share the model we have built with others. No matter how many models we create, if they remain offline, very few people will be able to see what we are achieving. That's why we should deploy our models, so that anyone can play with them through a nice User Interface (UI). For this system, we build a single page web application with Flask as the UI of our system.

Flask is a micro web framework written in Python. It is categorized as a microframework as it does not require any specific tools or libraries. It does not have a database abstraction layer, form validation or any other component where existing third-party libraries provide common features. The app accepts the input from person and gives the prediction regarding breast cancer as the output.

VI. CONCLUSION

In this project, CNN was used to classify cancer disease and implemented for Kaggle image dataset Then, the obtained classification accuracies were compared with each other.

The role of the classifier is crucial in the healthcare industry so that the results can be used for predicting the treatment which can be provided to patients.

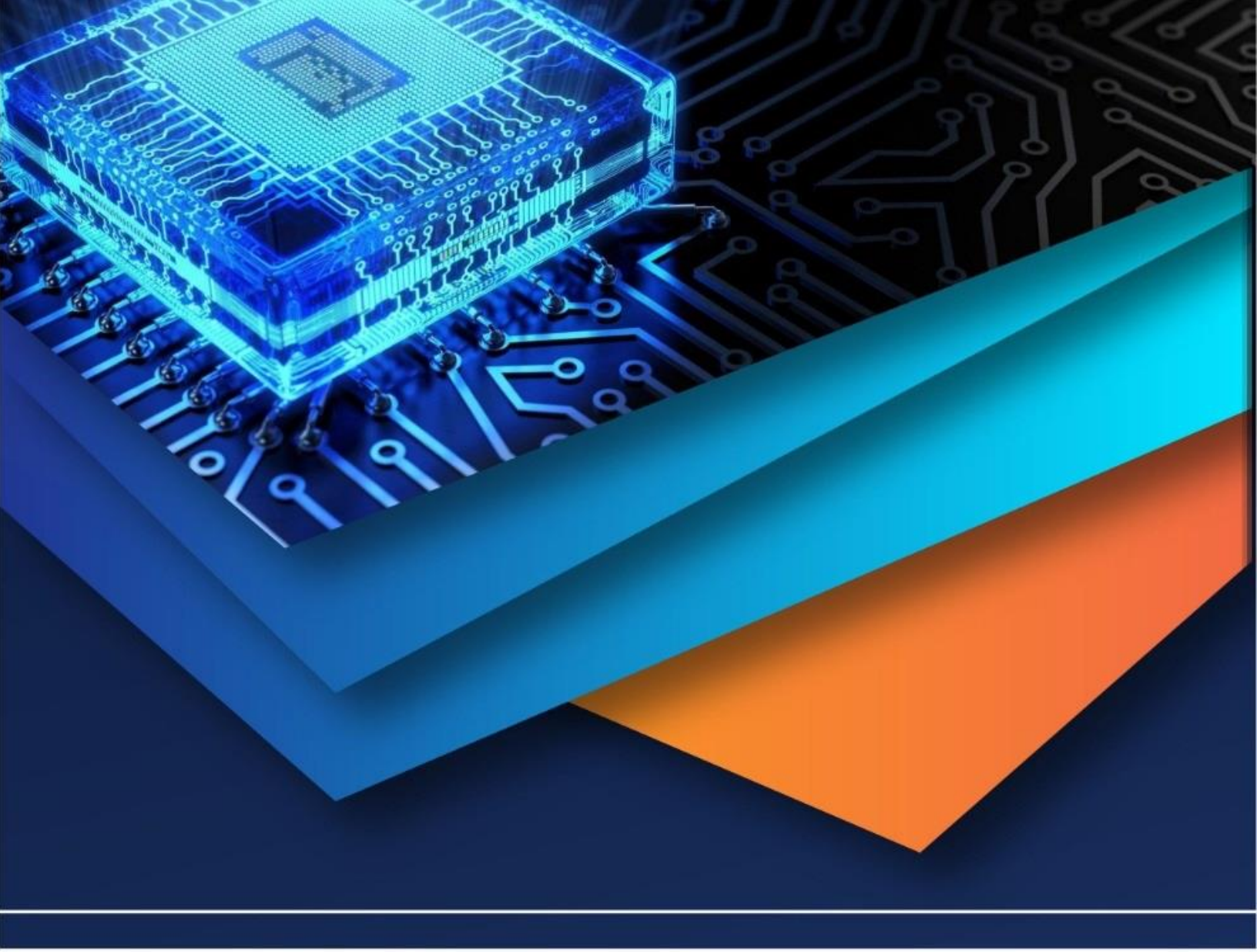
The existing techniques are studied and compared for finding efficient and accurate systems. It can be concluded that there is a huge scope for machine learning algorithms in predicting cancer diseases.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Alok Chauhan; Harshwardhan Kharpate; Yogesh Narekar; Sakshi Gulhane; Tanvi Virulkar; Yamini Hedau, 2021, " Breast Cancer Detection and Prediction using Machine Learning," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)
[2] Shams, Shayan, et al. "Deep generative breast cancer screening and diagnosis." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018.
[3] Guan, Shuyue, and Murray Loew. "Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks." Journal of Medical Imaging 6.3 (2019): 031411.
[4] Al-Dhabyani, Walid, et al. "Deep learning approaches for data augmentation and classification of breast masses using ultrasound images." Int. J. Adv. Comput. Sci. Appl. 10.5 (2019).
[5] M. Chen, Y. Ma, J. Song, C. Lai, B. Hu, "Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring,"ACM/Springer Mobile Networks and Applications' Vol. 21, No. 5, pp.825C845, 2016
[6] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," Health Affairs, vol. 33, no. 7, pp. 1123–1131, 2014
[7] D. Oliver, F. Daly, F. C. Martin, and M. E. McMurdo, "Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review," Age and aging, vol. 33, no. 2, pp. 122–130, 2004
[8] Quang H. Nguyen; Trang T.T. Do; Yijing Wang; Sin Swee Heng; Kelly Chen; Wei Hao, 2019, "Breast Cancer Prediction using Feature Selection and Ensemble Voting", International Conference on System Science and Engineering (ICSSE)
[9] Vinayak A. Telsang; Kavyashree Hegde, 2020, "Breast Cancer Prediction Analysis using Machine Learning Algorithms", International Conference on Communication, Computing and Industry 4.0 (C2I4), IEEE.
[10] Anuj Mangal; Vinod Jain, 2021, "Prediction of Breast Cancer using Machine Learning Algorithms", Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), IEEE.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)