iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** II **Month of publication:** February 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.40204

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 10 Issue II Feb 2022- Available at www.ijraset.com*

# Breast Cancer Detection with Machine Learning

Manav Mangukiya[1], Anuj Vaghani[2], Meet Savani[3]
[1]*Silver Oak College Of engineering & Technology, Ahmedabad, Gujarat, India*
[2]*Sarvajanik College Of engineering & Technology, Surat, Gujarat, India*
[3]*SAL College Of engineering, Ahmedabad, Gujarat, India*

***Abstract:*** ***Breast cancer is one of the leading cause for the death of women. In women Breast cancer is treated as the most significant issue. According to statistics released by the International Agency for Research on Cancer (IARC) in December 2020, Breast cancer has now overtaken lung cancer as the most commonly diagnosed cancer in women worldwide. Early diagnosis of this helps to prevent the cancer. If breast cancer is detected in early stage, then Survival rate is very high. Machine Learning methods are effective ways to classify data. Especially in the medical field, where those methods are widely used in diagnosis and analysis for decision making. In this paper, Data Visualization and performance comparisons between different machine learning algorithms: Support Vector Machine (SVM), Decision Tree, Naive Bayes (NB), K Nearest Neighbours (k-NN), Adaboost, XGboost and Random Forest conducted on Wisconsin breast cancer Dataset. The main objective is to evaluate the accuracy in the classification of data in terms of efficiency and effectiveness of each algorithm in terms of accuracy, precision, sensitivity and specificity. Our aim is to review various Techniques To detect early, efficiently and accurately Using Machine Learning. Experimental results show that XGboost offers the highest accuracy (98.24%) with the lowest error rate.***
***Keywords: Breast Cancer, Machine Learning, Wisconsin, Algorithms, Detection***

## I. INTRODUCTION

Total number of women dying in 2021 is approximately 963,000, according to the World Health Organization (WHO), Still, the organization predicts that the number could reach 2.9 million globally. Breast cancer can occur in women and rarely in men. The ICMR (Indian Council of Medical Research) recently published a report which stated that in 2020 the total number of new cancer cases is expected to be about 17.3 lakhs. An Indian woman is diagnosed with breast cancer in every four minutes. Breast cancer is a disease that occurs but when a woman or a man is aware of this symptom, it immediately goes beyond its original stage. Breast cancer is a common and dangerous disease in women, cancer is the creation of abnormal cells that come into these cells genetically and mutated. Spreads throughout the body, leading to death in diagnosis and treatment. There are two types of breast cancer, Malignant and Benign. The first is classified as harmful has the ability to infect other organs and is cancerous, Benign is classified as non-cancerous. This disease infects the women's chest and specifically glands and milk ducts, the spread of breast cancer to other organs is frequent and could be through the bloodstream. Different techniques are used to capture breast cancer such as Ultrasound Sonography, Computerized Thermography, Biopsy (Histological images).

Machine learning and Data mining techniques are straightforward and effective ways to understand and predict data. Radiologist examines and analyses himself and then he / she decides the result after participating with other experts. This process takes time and the results depend on the knowledge and experience of the staff. In addition, experts are not available in every field of the world. Therefore, the research community proposed automatic A system called CAD (Computer-Aided Diagnosis) for better classification of tumours, accurate results and faster Implementation without the need for radiologists or specialists. Machine learning algorithms (MLs) are indicated as one Option of human vision and experience to make final decisions with high accuracy.

Cancer in women always has a huge incidence rate and mortality rate. Breast cancer alone is estimated to account for 25% of all new cancer diagnoses worldwide and 15% of cancer deaths in women worldwide, according to the latest cancer statistics. Every 1 in 8 Women in USA develop breast cancer in her lifetime. In case of any sign or symptom, people usually visit a doctor immediately, who may refer you to an oncologist for help. An oncologist can diagnose breast cancer by: Examining the patient's medical history thoroughly, examining both breasts, and even checking for swelling or hardening of any lymph nodes in the armpits. Here in this project, we have used the Wisconsin Breast Cancer Dataset (WBCD) and with the dataset we have used machine learning algorithms to predict whether a patient has breast cancer. This paper compares the classification algorithms using an assembled approach suitable for demonstration and direct interpretation of their results. We are using the XGboost classified approach to compare the other classification algorithms and have analysed the accuracy of each classification of the best fit for breast cancer prediction.

©IJRASET: All Rights are Reserved 

## II. LITERATURE SURVEY

*A.* Turgut Machine learning procedure compared with SVM, KNN, DT, Logistic Regression, Random Forest, ADA Boost. In this various method checked and conclude that highest efficiency is 89% of random forest.

*B.* Narasingarao.M presents a survey of the work conducted to detect breast cancer using with different algorithm and conclude the efficiency of algorithm.

*C.* Junaid Ahmed achieved 84.21% accuracy by using Adaptive Reasoning Theory, the Wisconsin data set was used, that contains 569 rows of data, and also contains 32 attributes.

*D.* Nithya [13] applied the three categorizing methods such as Decision Tree, k-Nearest Neighbour, and Naïve Bayes for the different datasets. The authors also inspect the evaluation metrics of error rate. The implementation was focused on a type of attribute of a dataset.

*E.* Shilpa M and C. Nandini [19] implemented the algorithm using python and tested the same using dataset and achieved an accuracy of 94.74 and also reduces the time taken.

*F.* Hafizah [2] compared SVM and ANN using four different datasets of breast cancer. The researchers have demonstrated that SVM was better than ANN in performance and result.

*G.* S. Gc [1] worked on extracting features including variance, range, and compactness. They used SVM classification to analyse the performance. Their findings showed the highest variance of 95% and compactness 86%. According to their results, SVM can be considered as an appropriate method for Breast Cancer Prediction.

## III. METHODOLOGY

### *A. Dataset Description*

We have obtained Breast Cancer Wisconsin (Diagnostic) Dataset from Kaggle. Here 569 Patient's Data Was used for analysis, each instances have 32 Attributes with Diagnosis and Features. Each instance has a parameter of the cancerous non-cancerous cells and we will predict the cancer just by the input of features.

The values of features is in Numeric Format. The 'Target' means the patient Who is having Whether 'Benign' or 'Malignant' Cancer state. Benign means the patient is not having Cancer and Malignant means the patient is having Cancer.

TABLE I
Type of Patients

| Patient Type | Target |
|---|---|
| Benign | 1 |
| Malignant | 0 |

### *B. Data Visualization*

We are going to Visualize our Numeric data with Respect to Two categories 1) Benign 0) Malignant.

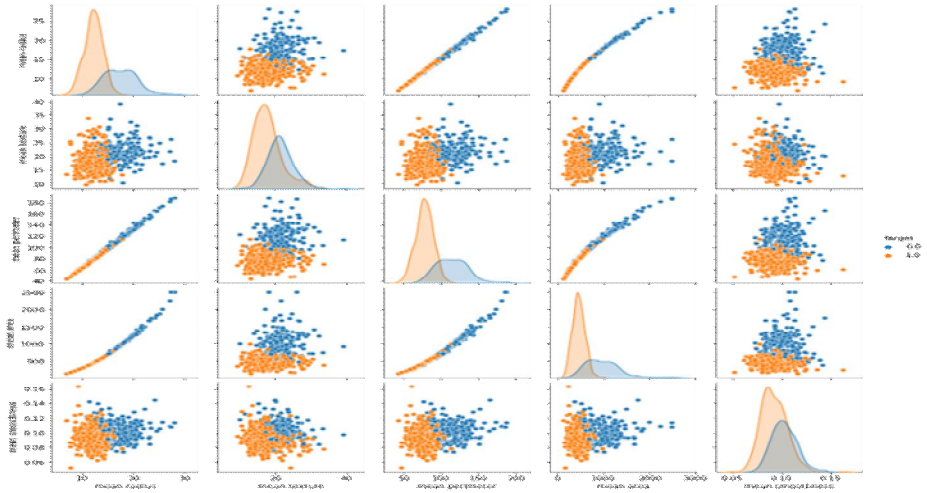


Fig. 1 Pairplot Of Features[mean_radius,mean_texture,mean_perimeter,mean_area,mean_smoothness]

### *C. Section Headings*

We used Google Colab and Microsoft Visual Code as a Coding platform and get a prediction output from the Flask in Local Server. Our Methods Includes Supervised Learning Algorithms and Classification Techniques like Support Vector Classifier (SVM), Random Forest, Naïve Bayes, Decision Tree, KNN, Adaboost and XGboost. Dataset contains features which highly vary in units and magnitudes. So, it is required to bring all features to the same level of magnitudes. We did that by using Standard Scaling in SKLearn.

Model selection is the most important step in Machine Learning. Machine Learning algorithms can be classified as: supervised learning and unsupervised learning. For Our project, we only need supervised learning. We used all Methodologies to Predict the result and Noted their Accuracy.

TABLE II
Comparison Between Techniques

| Techniques | Accuracy Without Standard scale | Accuracy With Standard scale |
|---|---|---|
| SVM | 57.89% | 96.49% |
| KNN | 93.85% | 57.89% |
| Random Forest | 97.36% | 75.43% |
| Decision Tree | 94.73% | 75.43% |
| Naïve Bayes | 94.73% | 93.85% |
| Adaboost | 94.73% | 94.73% |
| XGboost | 98.24% | 98.24% |

### *D. Confusion Matrix and Accuracy*

Confusion Matrix is used for evaluating the performance of a classification model. The Matrix compares the actual target values with predicted values by the machine learning model. It shows the ways in which your classification model gets confused when it makes predictions.

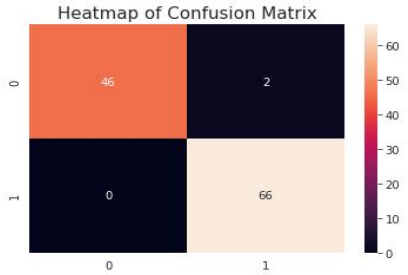



Fig. 2 Confusion Matrix

**Accuracy** Is given by:

**Accuracy** = (TP+TN)/(TP+TN+FP+FN)

= (46+66)/(46+66+0+2)*100 = 98.24

Where TP= True Positive , TN= True Negative, FP= False Positive, FN=False Negative

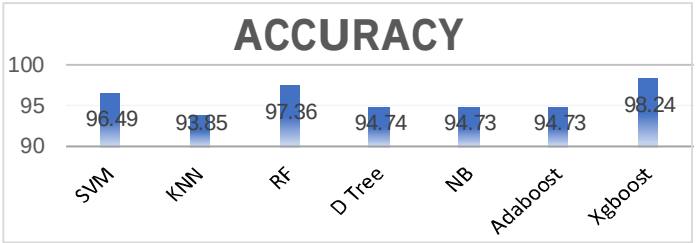



Fig. 3 Accuracy Comparison

## IV. PROPOSED SYSTEM ARCHITECTURE

As Shown in diagram, we first Uploaded dataset From Wisconsin Breast Cancer Dataset. After that We did Preprocessing to the data and applied Machine Learning Models, which is used in this project to predict Breast cancer.

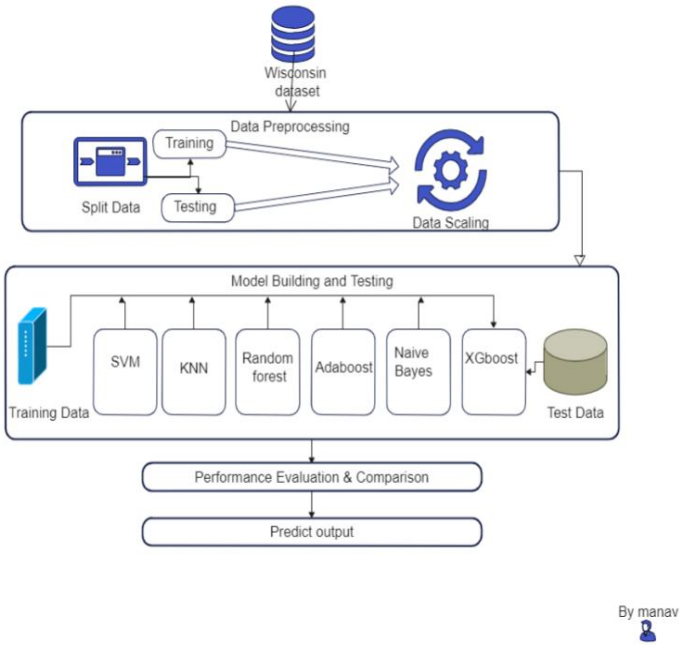



Fig.4 System architecture

## V. CONCLUSION & FUTURE WORK

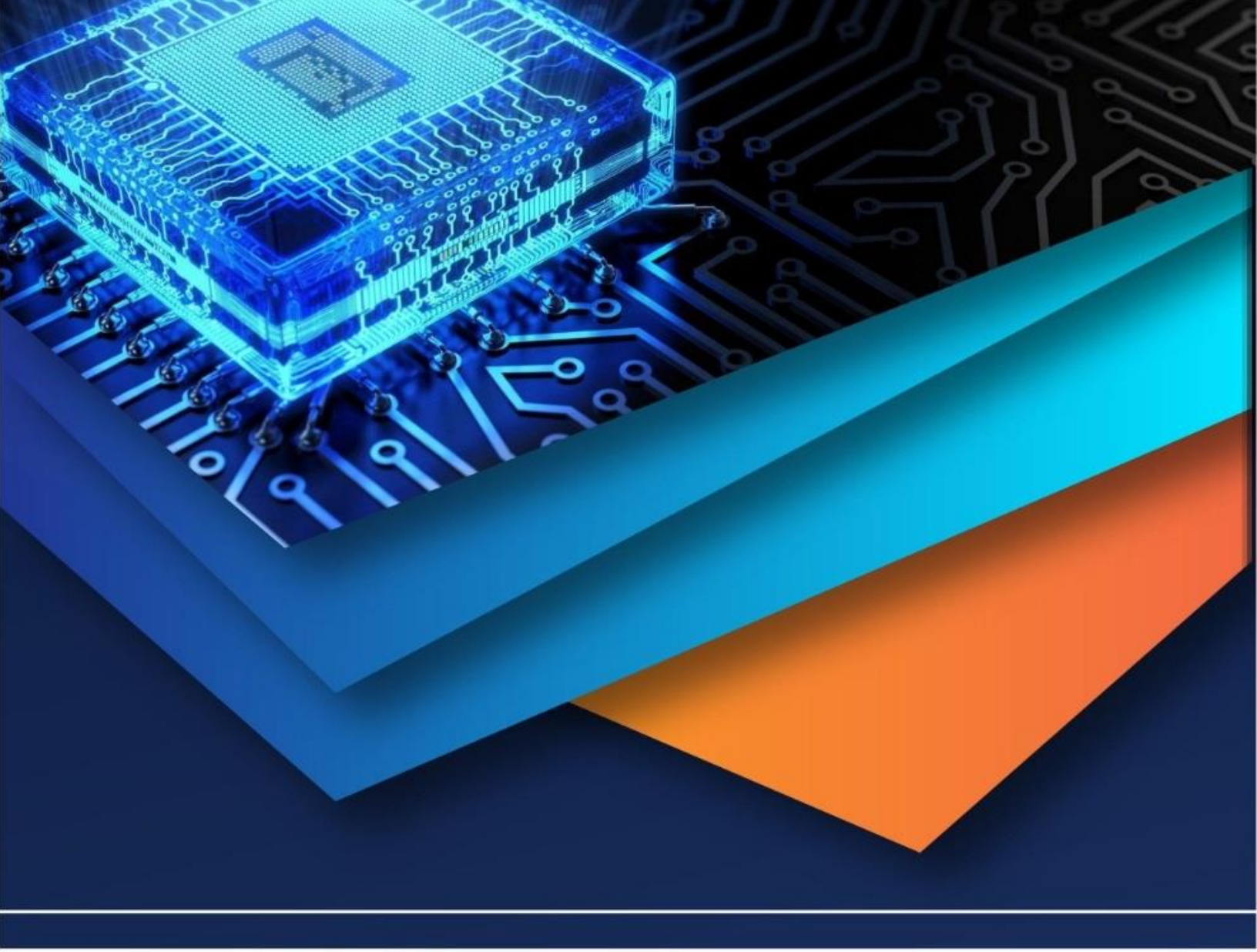
This paper examined different machine learning techniques for detection of breast cancer. The objectives of our study were to analyse the Wisconsin breast cancer dataset by visualizing and evaluating Machine Learning Predictions.With this research paper we can see that among Naïve Bayes, Support Vector Machine, Adaboost, Random Forest Classifier, KNN, Decision Tree, XGboost etc. We concluded that XGboost is the most accurate algorithm for best accurate result for detection of breast cancer with the efficiency of 98.24%. However, it is required that before running the algorithm, the dataset must be pre-processed. In future, we like to add larger dataset and check the efficiency and scalability of algorithm.

## REFERENCES

[1] S. Gc, R. Kasaudhan, T. K. Heo, and H.D. Choi, "Variability Measurement for Breast Cancer Classification Mammographic adaptive and convergent systems (RACS), Prague, Czech Republic, 2015, pp. 177–182.
[2] S. Hafizah, S. Ahmad, R. Sallehuddin, and N. Azizah, "Cancer Detection Using Artificial Neural Network and Support Vector Machine: A Comparative Study," J. Teknol, vol. 65, pp. 73–81, 2013.
[3] A. T. Azar, and S. A. El-Said, "Performance analysis of support vector Neural Compute. Appl., vol. 24, no. 5, pp. 1163–1177, 2014.
[4] machines classifiers in breast cancer mammography recognition," Neural Comput. Appl., vol. 24, no. 5, pp. 1163–1177, 2014.
[5] C. Deng, and M. Perkowski, "A Novel Weighted Hierarchical Adaptive Voting Ensemble Machine Learning Method for Breast Cancer 2015.
[6] Z. Jiang, and W. Xu, "Classification of benign and malignant breast cancer based on DWI texture features," ICBCI 2017 Proceedings of the Iinternational Conference on Bioinformatics and Computational Intelligence 2017.
[7] R. Jegadeeshwaran and V. Sugumaran (2013) Comparative study of decision tree classifier and best first tree classifier for fault diagnosis of automobile hydraulic brake system using statistical features, Measurement, vol.46, pp.3247–3260.
[8] Ajith Abraham (2005), Artificial neural networks, Nature & scope of AI techniques, vol.2, pp.901-908.
[9] Jennifer Listgarten, Sambasivarao Damaraju, Brett Poulin, Lillian Cook, Jennifer DuFour, Adrian Driga, John Mackey, David Wishart, Russ Greiner and BrentZanke (2004), Predictive Models for Breast Cancer Susceptibility from Multiple Single Nucleotide Polymorphisms, Clinical Cancer Research, vol.10, pp.2725- 2737.
[10] Jaree Thongkam, Guandong Xu and Yanchun Sang (2008), Breast cancer survivability via AdaBoost algorithms, Health data and knowledge management, vol.80.
[11] V. Sugumaran, V. Muralidharan and K.I. Ramachandran (2007), Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing, Mechanical Systems and Signal Processing, vol.21, pp.930-942.
[12] Hui-Ling Chen, Bo Yang, Jie Liuand Da-You Liu (2011) A support vector machine classifier with rough set- based feature selection for breast cancer diagnosis, Expert Systems with Applications, vol.38, pp.9014-9022.

[13] Tüba Kiyanand Tülay Yildirim (2004), Breast cancer diagnosis using statistical neural networks, Journal of electrical & electronics engineering, vol.4, pp.1149-1153.
[14] B. Nithya, V. Ilango, 2017, "Relative Analysis of categorization Methods in R Environment with two Different Datasets.", Intl J Scientific Research and Computer Science, Engineering and Information Technology (IJSRCSEIT), vol 2, Issue 6, ISSN: 2456- 3307.
[15] M. Shahbaz, S. Faruq, M. Shahen, and S. A. Masood, „„Cancer detection using data mining technology, Life Sci. J., vol. 9, no. 1, pp. 308–313, 2012.
[16] Pranay Shah, Rahul Deshpande, Nikhil Rao, Breast Cancer Detection System, (IRJET), Volume: 07 Issue: 05 | May 2020.
[17] Ajay Kumar, R. Sushil, A. K. Tiwari, Comparative Study of Classification Techniques for Breast Cancer Diagnosis, Vol.-7, Issue-1, Jan 2019.
[18] Vinoothna Manohar Botcha, Bhanu Prakash Kolla, Predicting Breast Cancer using Modern Data Science Methodology, ISSN: 2278-3075, Volume-8 Issue-10, August 2019.
[19] Sivapriya J, Aravind Kumar V, Siddarth Sai S, Sriram S, Breast Cancer Prediction using Machine Learning, ISSN: 2277-3878, Volume-8 Issue-4, November 2019.
[20] Shilpa M, C. Nandini "Breast Cancer Diagnosis and Prediction Using Machine Learning Algorithm" International Journal of Science and Research (IJSR) Volume 9 Issue 4, April 2020.
[21] Maria Mohammad Yousef, Big data analytics in healthcare: A review paper, International Journal of Computer Science & Information Technology (IJCSIT) Vol 13, No 2, April 2021.

©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)