International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6174
***Article Received**: 20 December 2022 **Revised**: 20 January 2023 **Accepted**: 27 January 2023*

Analysis and Classification of Breast Cancer Disease Via Different Datasets and Classifier Models

Ravi Kumar Barwal[1], Dr Neeraj Raheja[2], B R Mohan[3], Yamuna U[4], Sai Sudha Gadde[5], Dr. Madhu Patil[6]
[1]Research Scholar,
CSE Department, Maharishi Markandeshwar (Deemed to be University), Mullana, India
emails: barwal606@gmail.com[1]
[2]Associate Professor[2]
CSE Department, Maharishi Markandeshwar (Deemed to be University), Mullana, India
neeraj_raheja2003@mmumullana.org[2]

[3]Associate Professor,
Computer Science and Engineering, East West Institute of Technology, Bangalore
mohan.bangalore77@gmail.com
[4]Assistant professor ,
Department Of Information Science and Engineering
SJB Institute of Technology
uyamuna@sjbit.edu.in
[5]Assistant Professor
KoneruLakshmaiah Education Foundation, Vaddeswaram, A.P., India
Saisudhagadde@gmail.com
[6],Professor
Department of Electronics Communication and Engineering
BGS College of Engineering and Technology
madhu.ee@bgscet.ac.in

Abstract- Nowadays, Tumour is one of the important reasons of human death worldwide, producing about 9.6 million people in 2018. BC (breast cancer) is the common reason for cancer deaths in females. BC is a type of cancer that can be treated when detected early. The main motive of this analysis is to detect cancer early in life using ML (machine learning) techniques. The features of the people included in the WDBC (Wisconsin diagnostic breast cancer) and Coimbra BC datasets were classified by SVOF-KNN, KNN, and Naïve Bayes techniques. The pre-processing data phase was applied to the datasets before classification. After the data pre-processing steps, three classification methods were applied to the data. Specificity and Sensitivity rates were used to calculate the success of the techniques. As an outcome of the BC diagnosis classification, the SVOF-KNN technique was found with a 91 percent specificity rate and 90 percent sensitivity rate. When the outcomes attained from feature extraction and selection are calculated. It is seen that feature extraction, selection, and data pre-processing techniques improve the specificity and sensitivity rate of the detection system.

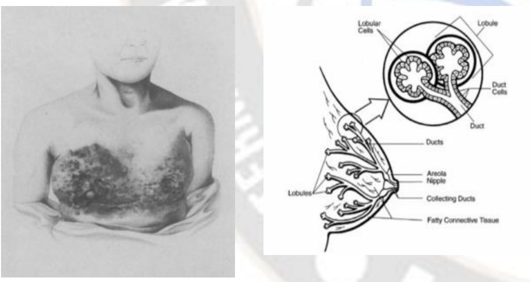
Keywords- Breast cancer detection,Wisconsin diagnostic breast cancer, Coimbra BC dataset, SVOF-KNN, KNN, Naïve Bayes.

I. INTRODUCTION

According to Bill Roth, the morphologic investigation of cancer was lately increased reliability. It provided better certainty in the clinical symptoms expanded with age. The medical experimental and anatomical features have overlapped fairly precisely through currently. Several experts tried to construct a classification, and various countries were confused and helpless to rise where a strong opinion was explained in detail seriously required. Bill Roth exasperated (tried) to laid items in the classification with the patient individually four types of breast cancer. As recommendatives of individuals who fixed the expressions or terms were developed broadly called, he selected his prototype at the *Viennese chair Franz Schuh*(1804 to 65) and *Sir John Birkett* (1815 to 1904). The processor of operation at *Louisville and Jefferson* Medical College in the US and Velpeau in Paris. The Billroth was an enormous measurable of cancer records statistically analyzed with his assistant Alexander von Win-water (1848 to 1917). An expert developed it in operation in *Liege, Belgium*, in 1878. According to the diagnosis, it seemed that the initial of Billroth's four sets. *Medullary carcinoma* was a faster course, and recurrent existed in young females aged between 35-40 years old. They initially appear as an infiltrative in the interval that extended immediately and before continuously taking the more significant part of breast cancer. Around 6-8 months, there was a widening of the axillary nodes monitored by the exterior of the supraclavicular credits. The

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6174
***Article Received**: 20 December 2022 **Revised**: 20 January 2023 **Accepted**: 27 January 2023*

condition of sustenance or diet remains better for a long though. The metastases grow in the pleura based on the unhealthy side in the liver and the long bones, not repeatedly, in the backbones. The patients are conveniently in excessive agony, which expiry will release them after 2-3 years at most [1]. Throughout the world, every year, breast cancer disease affects around 1.7 million females. BC is one of the riskiest diseases among other kinds of cancer, which is a widespread reason for death [2]. According to the report of the *American Cancer Society* in 2019, around 2,68,600 latest cases were identified as aggressive breast cancer disease patients. Further novel predictions of approximately 62,930 cases were found regarding in-situ breast cancer disease, with about 41,760 predictable death rates because of breast cancer. A recent breast cancer analysis is critical to increasing the number of survivors. The higher expensive cost of breast cancer diagnosis and extraordinary illness has motivated scholars to discover solutions to build further accurate models for cancer diagnosis [3]. The breast comprises two major forms of muscles (tissues), such as epithelial (glandular) and stromal (supportive) muscles. The epithelial tissues stock the milk-producing lobules or glands and the milk passages, whereas the stromal muscles of the breast. It is also composed of lymphatic tissues and immune structure muscles that eliminate cellular liquids and unwanted [4]. Fig. 1(i) and (ii), represent the primary structure of the breast.



(i) (ii)

Fig. 1(i) Cancer Disease and (ii)General structure of Breast [4]

This canceris a widespread disease of cancer that primarily exists in women and is the second most dangerous typical reason for cancer disease death in females in the US. It is (breast cancer) states to tumours or cancers initiated from the breast muscle and, most frequently, from the inside coating of milk lobules or ducts that provide the ducts using milk. It includes 10.4 percent of whole tumour rates among females, production it the 2nd public kind of non-tissue after lung tumour and the 5th widespread reason for death with cancer [4]. Several types of breast cancers exist, such as;

- Non-invasive (NIBC)
- DCIS
- LCIS
- IDC, and
- Invasive BC.

NIBC types of tissues are restricted to the vessels and do not enter near oily and connective cells of the breast. Ductal carcinoma in situ (DCIS) is another category of BC and is considered an augmented breast cancer hazard indicator. The invasive breast cancer tissues break down over the vessel, and lobular partition occupies the nearby fatty and associative cells of the breast. Infiltrating ductal carcinoma (IDC) is acategory of BC. It initializes in the breast milk lobular tissues and enters the duct partition, occupying the breast's fatty tissue and probably other body areas. Lobular carcinoma in situ (LCIS) is another form of BC. The word in situ denotes cancer that does not range preceding the region in which it is established initially. It is a severe growth in the numerous tissues in the breast lobules. Infiltrating lobular carcinoma (ILC) initiates in the breast's lobules and then frequently metastasizes to other areas of the human body.

Common symptoms or signs of BC, such as a lump. Regular breast self-examination is an effective method to familiarize the breast surface, cyclic variations, dimensions, and skin situation. The most alerting sign is swelling and mass in the breast and swelling in the lymph nodes, nipple release, aching in the nipple, retracted nipple, rough skin on the nipple, the insistent ache of the breast, and rare breast discomfort [4]. Genetic, hormonal, family, lifestyle, and environmental issues are responsible for breast cancer. Several factors raise a female's risk of breast cancer, called the genetic tendency of around five percent to ten percent of breast cancer [5] by the BRCA-1 exporters, and about 45% of breast cancer-2 (BRCA-2) holds [6] [7]. Females of *Ashkenazi Jewish* descent are called to be at higher threat of BRCA alternation when they also have great degrees for other actionable changes [8]. Another fewer typical gene changes consist of TP-53 and CHEK-2, such as the Li-Fraumeni set of symptoms. The CDHI includes genetic long-winded digestive cancer, STK-11 such as Peutz Jeghers genetic, PALB-2 interacts through BRCA-2 and the ATM such as ataxia-telangiectasia genetic factor. The chromosomal difference in stromal proteins decorin (SPD) [9] and the lumican using BC. Researchers in dual case-control revisions. The stroma is the helpful model of a biological cell in the breast involving different proteins like proteoglycans, decorin (DCN), and lumican (LUM). Another face of decorin and lumican is related to BC [10].

In 2020, experts combined external machine learning (ELM) through the Moth flame (MF) optimization method, a developed meta-heuristic to turn the EML model performance metrics. These performance metrics are such as values and hidden nodes to These performance metrics are such as values and hidden nodes to solve the issue of the challenging situation in the unseen layer of the model. It also utilized PCA and LDA for feature extraction and decreasing computational delay [11][12]. In 2019, an approach for BC classification with feature ensemble learning (EL) was established. They constructed a loaded sparse autoencoder (LSA) and softmax regression (SR) [13] to analyze benign and malignant tissues [14].

ML is a sub-set of implements used to form and design methods that predict and classify. ML depends on different phases: (i) medical data collection, (ii) medical data pre-processing, (iii) feature extraction (FE), and detection (training and testing module).

This analysisdescribed different ML methods such as SVOF-KNN, KNN, and NB (naïve bayes) classifiers. The main motive is to determine whether a patient has been infected,or non-infected. The SVOF-KNN is a proposed model which is recommended the BC patients. This method has improved performance metrics and compared them with existing ones.

The lasting Sections of this article are managed as trails: Section 2 defines a review of literature, which means the existing proposed methods, issues, datasets, problems, and tools with parameters. Section 3 defines the proposed work that implements usingvarious datasets and classifiers, and Section 4 represents the outcome analysis. Section 5 determines the primary analysis of the research work and further improvements.

II. RELATED WORKS

Hanan Aljuaid et al. (2022) [15] described several developed and under-progress nations worldwide suffering from cancer as a severe disease. Specifically, the ratio of breast cancer primarily exists in women, was raised day by day moderately. The main reason for cancer's growth was lack of knowledge and unidentified at the initial states. The proper initial analysis effectively detects cancer disease that reduces the chance of breast cancer disease. Additionally, the classification of cancer in the initial stage also eliminated the growth of cancer disease, with the help of healthcare image detection and computer vision, accelerated and automated cancer analysis. For large volumes of the database of healthcare images and convolution neural network (CNN) was the most wide-ranging approach for detecting or classifying cancer disease. The authors represented novel automated diagnosis techniques for breast cancer organization such as binary and multiple class with the hybridization of deep neural networks (DNNs) such as ResNet-18, Inception-V3Net, and Shuffle-Net, and transfer learning (TL) with openly available datasets. The proposed technique achieved better outcomes as the accuracy parameters for binary and multiple classes separately. Laith Alzubaibi et al. (2020) [16]described breast cancer disease as a major issue in women's mortality. An initial cancer diagnosis was to reduce breast cancer's chances and decrease the death rate. Using a computerized diagnosis cancer prediction model that improved efficiency at low cost. The traditional breast cancer prediction model did not provide relevant results due to a lack of technology and was handcraft-based. Still, medical breast cancer data were complex in texture. Existing DLL models were an alternative result for analysis and eliminated the complex traditional techniques. The most significant challenge was training data limitation and performance improvement. The authors proposed a method using deep CNN for breast cancer detection. The authors utilized TL in dual forms, such as the trained prosed method initially on a similar dataset, then on the aimed dataset employed. The proposed hybrid method works in different parts for breast cancer detection. Four different modules were considered as invasive carcinoma, benign tumour, in-situ carcinoma, and typical tissues. The proposed hybrid deep CNN method achieved 90.5 % accuracy for patch-based classification and 97.4% accuracy using the microscope ICIAR-2018 dataset. Haifeng Wang et al. (2017) [17] described an SVM-based collaborative MLtechnique for the analysis of breast cancer. Breast cancer, a wide-ranging performed disease, plays a dangerous role in diagnosis labeling policies frequently interrelated to patient safety. Currently, several classification systems exist in the data-mining area. The classification areas used for breast cancer disease were improved based on patients' health records. But the accuracy of every method was based on different classical formations like input attribute form and model parameters. Breast cancer disease was an analysis thatutilized the SVM-based collaborative learning method to eliminate the analysis modifications and improve the diagnostic accuracy. Around 12 SVMs based on the constructed weighted AU, the receiver operative characteristics (ROC) method were composed. The evaluation of the implemented hybrid technique was calculated using the WDBC and surveillance epidemiologic and end resolution (SEER) dataset. The investigational outcomes of the implemented hybrid method reached adequate accuracy using considerably fewer changes for breast cancer diagnosis. Then it was compared with five other collaborative approaches and two frequently used models as adaptive and bagging classification tree methods. The proposed hybrid method reached an accuracy of 97.89%. Deepti Sharma et al. (2022) [18] described cancer forecast as a significant and challenging factor for surgeons

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6174
***Article Received**: 20 December 2022 **Revised**: 20 January 2023 **Accepted**: 27 January 2023*

and experts. BC detection at an initialphase helped in the diagnosis and prediction timely, and several investigators proposed different techniques for cancer disease initial forecast. The experts proposed a method using the ensemble learning-based NN and an additional classification tree for breast cancer disease prediction. This classification was divided into two parts: cancerous, for example, malignant, and non-cancerous, for example, benign. The authors utilized UCI and WDBC datasets for this proposed method implementation. The experimental outcomes of the plannedtechnique achieved better performance parameters like accuracy, recall, precision, f-measure, etc. The developed method's forecast (prediction) effects reached an overall accuracy of 99.74 %. Furthermore, the proposed approach using NN and an additional tree overtakes other advanced classifiers in the form of different performance parameters. The proposed method recommended was proven effective and encouraging for breast cancer classification and was additionally represented using experimental simulated and observed outcomes and statistical evaluations. Marion Olubunmi Adebiyi et al. (2022) [19] described those cases detected in the final stage of BC. Cancer, considered the most dangerous disease, especially breast cancer, exists in women. Breast cancer was an unrestricted malignancy in females globally. However, mammography designed to detectBC does not usually exist in all clinics, and the time consumption in analysis and control ofBC increases the chance of the growth of the disease. Computerized diagnosis methods were utilized based on machine learning approaches (ML) toexpand breast cancer detection and reduce the mortality ratio. The authors proposed a way for an investigation to expand the accuracy of ML methods for breast cancer detection. The ML approaches permitted the taxonomy and estimate of breast cancer. The authors required ML approaches of random forest (RF) and SVM using the feature extraction (FE) approach of linear discriminant analysis (LDA) using the WDBC dataset. ML approaches permitted the taxonomy and forecast of cancer. The experimental outcomes of the proposed strategy achieved an accuracy of 96.4 % and 95.6% for SVM and RF with LDA. Nagwan Abdel Samee et al. (2022) [20] described a promising medical field, and the systematic unrestricted was concentrated on the artificial intelligence (AI) applications used for actual healthcare issues like as construction of computer-aided diagnosis (CAD) models designed forBC. The TL technique was considered a modern AI-based approach that permitted fast learning growth and recovered medical data diagnosis improvement. The DL classification for breast cancer broadly protected specific difficulties, such as the issues still in the proposed investigated method. The first issue was to improve the input data of the DL models by organizing pseudo-colour images as an alternative to single-input real grayscale image data. The IP methods were simultaneously utilized to reach the specific goal. The CNNs model was used to develop a more accurate. Other reflected issuesin moderating the numerous collinearity and topographies from DL models. The authors proposed a hybrid method with LR's composition and principle components analysis (PCA). This proposed method helped to choose the appropriate principal components (PCs). The proposed method achieved an accuracy of 98.60% (INbreast dataset) and 98.80% (mini MAIS dataset. Table 1 discusses the proposed methods, problems, gaps, performance metrics, and dataset in breast cancer classification.

TABLE 1. COMPARATIVE ANALYSIS OF THE DIFFERENT METHODS

AuthorName/ Year	**Proposed Technique**	**Problems/Gaps**	**Parameters**	**Dataset**
Hanan Aljuaid et al. (2022) [15]	Computer-Aided Diagnosis Technique	Lack of accuracy	Precision Sensitivity Specificity Accuracy	BreakHis dataset
Laith Alzubaibi et al. (2020) [16]	Hybrid DCNN Approach	Lack of training data.	Patch–wise and image-wise Accuracy	ICIR2018, Microscopy images
Haifeng Wang et al. (2017) [17]	The SVM-based weighted area under Curve (AUC) method	It is suitable for limited aspects of cancer disease detection.	Precision Sensitivity Specificity Accuracy	SEER WDBC
Deepti Sharma et al. (2022) [18]	The proposed feature ensemble learning based NN	Feature selection is a significant issue in the diagnosis of breast cancer.	Precision Sensitivity Specificity Accuracy F-measure MCC	WDBC UCI dataset

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6174
***Article Received**: 20 December 2022 **Revised**: 20 January 2023 **Accepted**: 27 January 2023*

			Accuracy	
Marion Olubunmi Adebiyi et al. (2022) [19]	SVM and RF-based proposed methods for breast cancer detection.	The problem of limited accuracy occurs in breast cancer detection.	Precision Sensitivity Specificity Accuracy F-measure	WDBC
Nagwan Abdel Samee et al. (2022) [20]	The hybrid method was proposed using PCA and LR.	The multiple collinearity issue occurs in the detection process.	-	INbreast dataset and mini MAIS dataset

III. PROPOSED WORK

A. Dataset 1 Breast Cancer Coimbra Dataset [21]

This dataset'sdefinite and algebraictraitson behalf of this proposed investigation were reached from the openly availableon the website. Table 2 comprehensively clarifiesthis dataset's traits [21]. It holds64 individual records through BC and 52Well patient data accessed based on the medical structures of the whole dataset of 116 occurrences. This UCI-accessibleMLsource was produced in 2018 [22]. The dataset's computationaltraits are autonomous metrics, and certain features depend on a metric.

High BMI is the leading risk factor for BC. Also, cancer growth is most probable when there is an increased level of resisting leptin and minimized adiponectin secretion. After menopause, women are at a higher risk of developing BC due to metabolic syndrome, including belly obesity and insulin conflict. HOMA-IT can be utilized to verify the patients' records by subdivision-medical insulin resistance. Permitting to the bc dataset, the relationship concerning insulin, glucose, resistance to bmi, and adiponectin [23] are the significant constraints for the development of bc.

TABLE 2. MEDICAL ATTRIBUTES OF THE BC COIMBRA DATASET [21]

Attribute	Description	Range/Units
Age	Patient Age year-wise	Range- 24-89
Body mass index	BMI (Body mass index) of the particular.	Units- kg/m^2 Range – 18-38
Glucose	The glucose level of the particular	Units- mg/dl
Insulin	Insulin level of the patient	Units - μg/ml
Leptin	Leptin level of patient	Units – ng/ml
Adiponectin	Adiponectin secretion level of the patient	Units- μg/ml
MCP-1	Monocyte Chemo-attractant Protein-1	Units- pg/dl
Classification	Healthy (1); patients (2)	-

B. Dataset 2 WISCONSIN Breast Cancer Dataset [21]

The BC dataset we utilized contributed to UCI (the University of California Irvine). This dataset has 11 attributes, and the initial one is the ID we will remove because ID is not a feature we want to feed into our classification. The nine standards are explained earlier in the BC classification phase, which means determining whether the cancer is BENIGN or malignant. The last feature gives a binary number “2” representing BENIGN cancer and “4” representing cancer. The set comprises 699 medical cases. The first breast cancer detection consistedof missing information for 16 observations, which limited the research database to 683 samples.

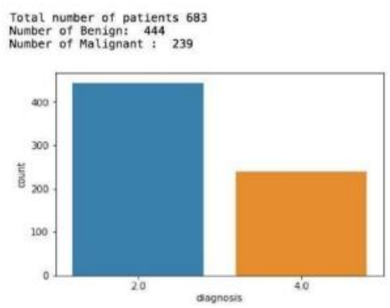


Fig 2. Wisconsin BC Datasets [21]

TABLE 3. WISCONSIN DATASET [26]

Attribute Information:
ID number
Diagnosis: M (malignant), and B (Benign)
Radius (distance mean)
Perimeter
Area
Smoothness
Compactness
Concavity
Concave Points
Symmetry
Fractal Dimension

Fig 2 represents that 444 means 65 percent of cancers are BENIGN cancers and 239 means 35 percent are malignant. Table 3 represents the dataset attributes information.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6174
***Article Received**: 20 December 2022 **Revised**: 20 January 2023 **Accepted**: 27 January 2023*

C. Classification SVOF-KNN Model

The classification structure SVOF-KNN is replicatedwiththe ML lending library in MATLAB 2018a to suggestforecast or classification methods reliable for reliable, precise standards. The proposed "singular value optimized Feature-based KNN" model is from the ML-based model. In this proposed model, the various neighbours pooled to compute an estimateareusedlowest than the centroid of K. In This proposed work, the singular value optimizedFeature-based KNN method is deployed for orderingor forecastinvestigation. The proposed method has been applied for approval: extracting the features and choosingdependablequalities between the healthcare databasequalities and raising the ML-based practice for the forecast.

D. Naïve Bayes (NB) Classifier [21]

NB classifier approach estimates that the classifier attribute is liberated and may not have a correlation among them. Numerous investigators analyzed that the supposition of independence may operate in all cases for which other methods are developed to improve the performance rate. The bayesian technique depends on the conditional possibility and maximum probability event. The bayesian algorithm is given as follows:

- Assume G be the train_set with X tuples where every tuple is demonstrated as 'L' dimension attribute vector O, where O= $\{O_1, O_2, \ldots\ldots O_k\}$.
- Assume 'b' classes z_1, $z_2, \ldots\ldots z_P$. Based on the Bayesian approach, a tuple T is related to class z_xOnly when it has a high uncertain possibility than another class z_yHere x ≠y.

$$b\ (z_x \mid \mathrm{Ti}>b(z_y \mid \mathrm{Ti})\ \text{and}$$

$$b\ (z_x \mid \mathrm{T} = \ (b(\ \mathrm{Ti}\ (z_x * b\ ((z_x\ ))\ /b(\mathrm{Ti}).$$

- When conditional class individuality is supposed.

$$b(O|\ (z_x\ )\ = \textstyle\prod_{j=1}^{k}\ b(O_K \mid (z_x)\ =\ b(O_1\ ) * bz(O_2\ ) * b(O_3\ ) \ldots\ldots\ldots\ldots * b(O_k\ ).$$

- Class (z_x is projected as the productivity class while

z (O| (z_x) * P(z_y) > (O | z_y), where 1 ≤ x1, y2≥ z and x1≠ y1.

E. KNN Classifier

It is a non–parametric technique that depends on the usage of space measurement. All the accessible cases may be placed in it, and when a new case arrives, it may classify based on the distance value [22]. The mathematical equation for K nearest neighbour algorithm is given by;

$$D= \sqrt{(x_2 - x_1)^2} + \sqrt{(y_2 - y_1)^2} \ldots..(i)$$

KNN influences the distance equation, whose main objective is searching for the nearest neighbour in the coordinate system. The distance between the two points arrives from Euclidian geometry and different dimensions. For example; the value,x_1 for the height, x_2 for the weight in equation (i) and fig 4 shows the categorize the number of classes (A,B).

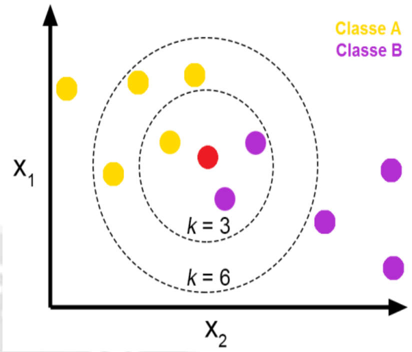


Fig 3. KNN classifies the number of Classes (A,B) [22]

IV. RESULT ANALYSIS

In this investigation, the SVOF-KNN methodreplications were approvedwith a singular value-optimized Feature-based KNN method. This forecastmethodologytasks the hazard of BC bythe proposed model. It orders the finest classification regarding the parameters in the form ofspecificity and sensitivity, etc.

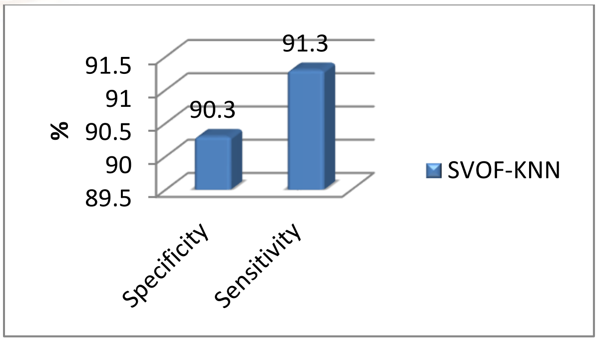
The comfortabledata of the BC Coimbra and Wisconsin BC datasets are accessed, and the BC evidence is pre-processed for FE (feature extraction), collection, and modules. Currently, these datasets are uploaded in the reading class for exemption three classifications.

TABLE 4. COMPARISON ANALYSIS

Methods	Specificity	Sensitivity
SVOF-KNN	90.3	91.30
KNN[24]	78.13	84.62
Naïve Bayes [23]	91.5	88.8

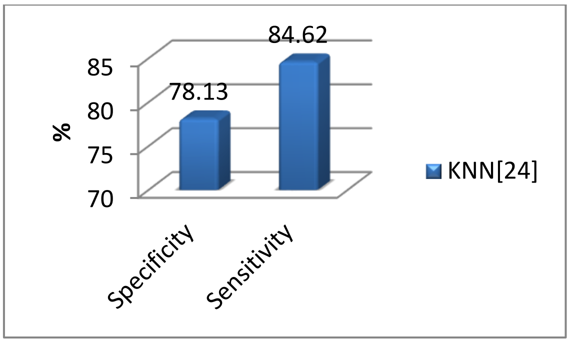
Table 2 discusses the specificity and sensitivity attained on the classification methods with BC Coimbra BC Wisconsis datasets and data visualization of SP and SN defined in fig 4(i), 4(ii), and 4(iii).

and

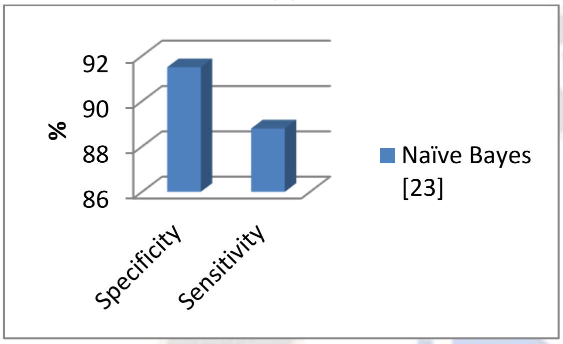


(i)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6174
***Article Received**: 20 December 2022 **Revised**: 20 January 2023 **Accepted**: 27 January 2023*



(ii)



(iii)

Fig 4. Performance metrics with different algorithms (i) SVOF-KNN, (ii) KNN[24], and (iii) Naïve Bayes [23]

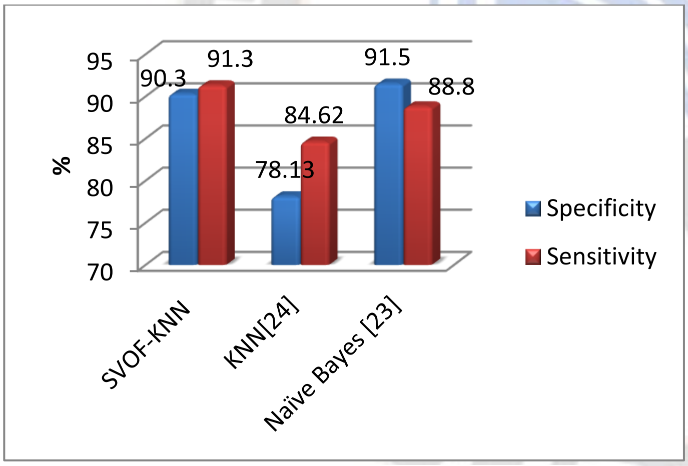
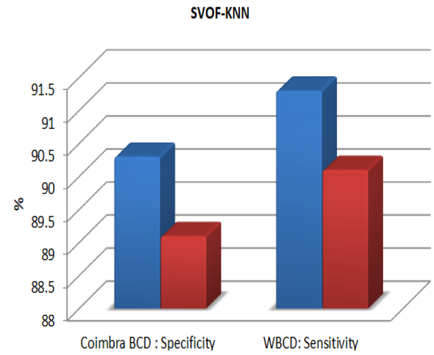


Fig 5. Comparison analysis with different methods: Specificity and Sensitivity rate

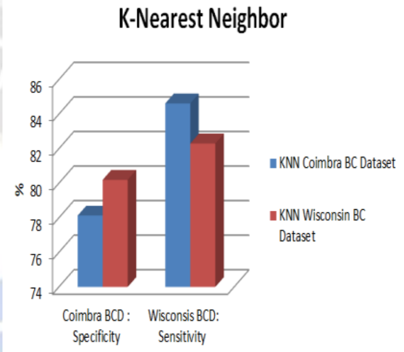
Fig 5 represents the comparison analysis with SVOF-KNN, KNN, and Naïve Bayes method with different parameters specificity, and sensitivity rates. The proposed SVOF-KNN model has achieved a higher SP and SN rate as compared to other methods.

TABLE 5. DATASET COMPARISON

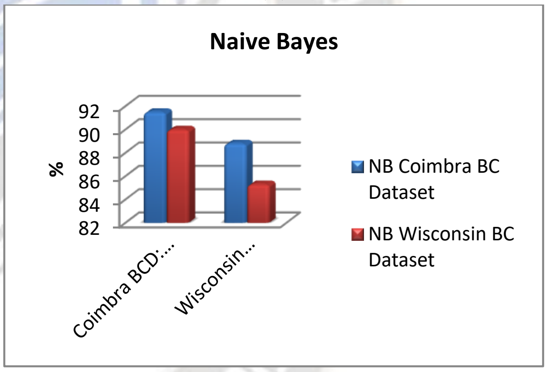
Dataset| Methods		Specificity	Sensitivity
SVOF-KNN	**Coimbra BC Dataset**	90.3	91.30
	Wisconsin BC Dataset	89.1	90.1
KNN	**Coimbra BC Dataset**	78.13	84.62
	Wisconsin BC Dataset	80.2	82.3
NB	**Coimbra BC Dataset**	91.5	88.8
	Wisconsin BC Dataset	90.0	85.3



(i)



(ii)



(iii)

Fig 6. Comparative Analysis with Coimbra BC dataset 1, and Wisconsin BC dataset 2 (i) SVOF-KNN (ii) KNN, and (iii) Naïve Bayes with SP and SN rates

Fig 6 represents the comparison analysis with different BC datasets, different parameters specificity, sensitivity rates, and methods. The proposed SVOF-KNN model has achieved a higher SP and SN rate as compared to other methods.

V. CONCLUSION AND FUTURE SCOPE

This research work utilized the WISCONSIN and COIMBRA BC datasets to analyze the most reliable and successful BC SVOF-KNN classification model. SVOF-KNN, KNN, and Naïve Bayes (NB) techniques were used in the classification and recommendation. For a general comparison of accomplishments between processes, it is seen that the Coimbra BC dataset with the SVOF-KNN

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6174
***Article Received**: 20 December 2022 **Revised**: 20 January 2023 **Accepted**: 27 January 2023*

method is the most reliable method, with a specificity value of 90.3 percent and a sensitivity of 91.30 percent. This is followed by the KNN and Naïve Bayes methods with an SP value of 78.1 and 84.6 percent. The performance parameter values of the methods with the maximum value for each technique are defined in table 5. For future analysis, we are consideringproposing a single chip-based NNs (neural networks) to diagnose the irregularities of EEG, heart rate, etc.

REFERENCES

[1]. De Moulin, D. (2012). *A short history of breast cancer*. Springer Science & Business Media.

[2]. Miller, K. D., Siegel, R. L., Lin, C. C., Mariotto, A. B., Kramer, J. L., Rowland, J. H., ... & Jemal, A. (2016). Cancer treatment and survivorship statistics, 2016. *CA: a cancer journal for clinicians*, *66*(4), 271-289.

[3]. Alzubaidi, L., Al-Shamma, O., Fadhel, M. A., Farhan, L., Zhang, J., & Duan, Y. (2020). Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model. *Electronics*, *9*(3), 445.

[4]. Sharma, G. N., Dave, R., Sanadya, J., Sharma, P., & Sharma, K. (2010). Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology & research*, *1*(2), 109.

[5]. Claus, E. B., Schildkraut, J. M., Thompson, W. D., & Risch, N. J. (1996). The genetic attributable risk of breast and ovarian cancer. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, *77*(11), 2318-2324.

[6]. Narod, S. A., & Foulkes, W. D. (2004). BRCA1 and BRCA2: 1994 and beyond. *Nature Reviews Cancer*, *4*(9), 665-676.

[7]. Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., ... & Breast Cancer Linkage Consortium. (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *The American Journal of Human Genetics*, *62*(3), 676-689.

[8]. Frey, M. K., Sandler, G., Sobolev, R., Kim, S. H., Chambers, R., Bassett, R. Y., ... & Blank, S. V. (2017). Multigene panels in Ashkenazi Jewish patients yield high rates of actionable mutations in multiple non-BRCA cancer-associated genes. *Gynecologic Oncology*, *146*(1), 123-128.

[9]. Monticciolo, D. L., Newell, M. S., Moy, L., Niell, B., Monsees, B., & Sickles, E. A. (2018). Breast cancer screening in women at higher-than-average risk: recommendations from the ACR. *Journal of the American College of Radiology*, *15*(3), 408-414.

[10]. Kelemen, L. E., Couch, F. J., Ahmed, S., Dunning, A. M., Pharoah, P. D., Easton, D. F., ... & Vachon, C. M. (2008). Genetic variation in stromal proteins decorin and lumican with breast cancer: investigations in two case-control studies. *Breast cancer research*, *10*(6), 1-11.

[11]. Muduli, D., Dash, R., & Majhi, B. (2020). Automated breast cancer detection in digital mammograms: A moth flame optimization based ELM approach. *Biomedical Signal Processing and Control*, *59*, 101912.

[12]. Allugunti, V. R. (2022). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *International Journal of Engineering in Computer Science*, *4*(1), 49-56.

[13]. Kadam, V. J., Jadhav, S. M., & Vijayakumar, K. (2019). Breast cancer diagnosis using feature ensemble learning based on stacked sparse autoencoders and softmax regression. *Journal of medical systems*, *43*(8), 1-11.

[14]. Aslam, M. A., & Cui, D. (2020, July). Breast cancer classification using deep convolutional neural network. In *Journal of Physics: Conference Series* (Vol. 1584, No. 1, p. 012005). IOP Publishing.

[15]. Aljuaid, H., Alturki, N., Alsubaie, N., Cavallaro, L., & Liotta, A. (2022). Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Computer Methods and Programs in Biomedicine*, *223*, 106951.

[16]. Alzubaidi, L., Al-Shamma, O., Fadhel, M. A., Farhan, L., Zhang, J., & Duan, Y. (2020). Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model. *Electronics*, *9*(3), 445.

[17]. Wang, H., Zheng, B., Yoon, S. W., & Ko, H. S. (2018). A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, *267*(2), 687-699.

[18]. Sharma, D., Kumar, R., & Jain, A. (2022). Breast cancer prediction based on neural networks and extra tree classifier using feature ensemble learning. *Measurement: Sensors*, *24*, 100560.

[19]. Adebiyi, M. O., Arowolo, M. O., Mshelia, M. D., & Olugbara, O. O. (2022). A Linear Discriminant Analysis and Classification Model for Breast Cancer Diagnosis. *Applied Sciences*, *12*(22), 11455.

[20]. Samee, N. A., Alhussan, A. A., Ghoneim, V. F., Atteia, G., Alkanhel, R., Al-Antari, M. A., & Kadah, Y. M. (2022). A Hybrid Deep Transfer Learning of CNN-Based LR-PCA for Breast Lesion Diagnosis via Medical Breast Mammograms. *Sensors*, *22*(13), 4938.

[21]. Amrane, M., Oukid, S., Gagaoua, I., & Ensari, T. (2018, April). Breast cancer classification using machine learning. In *2018 electric electronics, computer science, biomedical engineerings' meeting (EBBT)* (pp. 1-4). IEEE.

[22]. Hu, Li-Yu, Min-Wei Huang, Shih-Wen Ke, and Chih-Fong Tsai. "The distance function effect on k-nearest neighbor classification for medical datasets." *SpringerPlus* 5, no. 1 (2016): 1-9

[23]. Thirumalaikolundusubramanian, P. (2018). Comparison of Bayes classifiers for breast cancer classification. *Asian*

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6174
***Article Received**: 20 December 2022 **Revised**: 20 January 2023 **Accepted**: 27 January 2023*

Pacific journal of cancer prevention: APJCP, *19*(10), 2917.

[24]. Kanimozhi, G., Shanmugavadivu, P., & Rani, M. M. S. (2020). Machine Learning-Based Recommender System for Breast Cancer Prognosis. *Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries*, 121-140.

[25]. Borges, L. R. (1989). Analysis of the wisconsin breast cancer dataset and machine learning for breast cancer detection. *Group*, *1*(369), 15-19.