Power System Technology
ISSN:1000-3673

Received: 06-06-2024 *Revised: 15-07-2024* *Accepted: 28-08-2024*

Diagnosis of Breast Cancer Using Entropy Weight-Feature Selection and Ensemble Learning

Fatemeh Jafari[1], Hamidreza Ghafari[2*]

1. Department of Computer Engineering, Ferdows Branch, Islamic Azad University, Ferdows.Iran.
Email: fatemejafari.edu@gmail.com
2. Department of Computer Engineering, Ferdows Branch, Islamic Azad University, Ferdows.Iran (Corresponding author).
Email: hamidghaffaryj53@gmail.com

Abstract- The discovery of huge medical databases with the help of new computational tools has confirmed the existence of all diseases, including certain cancers. However, medical research now tends to look at diseases separately from each other, rather than focusing on their interactions. So far, various methods based on data mining and machine learning techniques have been proposed, which, despite their many applications in the field of diagnosis of breast cancer, still face the problem of insufficient accuracy, which has become a major challenge. Therefore, in this paper, for the diagnosis of breast cancer using the entropy-based feature selection algorithm, the weight-prominent feature is selected and then the ensemble learning system consisting of SVM, KNN, Adaboost and ID3 decision tree machine learning algorithms used. In this paper, the SEER dataset has been used. After simulating the proposed method, it was observed that the accuracy of the proposed method on average compared to other methods such as SVM, KNN, Adaboost and ID3 decision tree on breast cancer data has improved by about 11.5% compared to other methods.

Keywords: Entropy weight-feature selection, breast cancer, ensemble learning

1- INTRODUCTION

With the development of computer application in medical science, it has become possible for a large number of diseases to be detected early with the help of machine learning techniques. Cancer occurs when some cells in the body begin to grow and multiply abnormally. Cancer is said to have progressed when the gene for cell growth and proliferation was altered. Eventually, these mutations will become a mass through cell proliferation. If the cancer-transmitting gene is identified, significant progress has been made in predicting cancer. Diagnosing diseases such as cancer as soon as possible will be very helpful in treatment [1]. Cancer is a major problem in all countries of the world. Although accurate statistics are not yet available, it is estimated that the number of new cases of cancer in the United States by 2012 was about 2 million [2].

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

According to the British Cancer Research Center, in 2012, the number of new cases of cancer in the world was more than 14.1 million, of which more than 8.2 million were fatal. According to the latest figures from the National Cancer Institute, in the United States alone, the number of new cases of cancer in 2016 was more than 1.6 million, of which more than 35 percent died. Among these, breast and genital cancer (in women) and prostate (in men) account for about 30% of common cancers [3]. Given that most cancers are breast cancer in women, in this paper we focus on this cancer.

Breast cancer is one of the most common types of cancer in women, affecting about 10% of all women in their lifetime. Cancer tumors are divided into benign and malignant. Benign tumors only grow abnormally, but in very rare cases they can cause death. However, a number of benign breast masses can also increase the risk of breast cancer [4]. Also, in some women with a history of breast biopsy, the risk of breast cancer increases. Malignant tumors are very dangerous and are considered cancerous, but early detection of this type of cancer has been very effective in successful treatment [4]. One of the applications in breast cancer is to predict its recurrence using machine learning techniques. Machine learning can help doctors make decisions by reducing the positive results that are wrong and the negative results that are wrong [5]. New techniques, such as the discovery of knowledge from databases, which are done using machine learning techniques, are becoming more popular every day and have become a popular tool for medical researchers. In this way, researchers can identify patterns and relationships between a large number of variables and predict the results of a patient using data stored in a database.

Various techniques and methods have been proposed by various researchers to diagnose cancer. Each of these methods is based on a specific strategy. Fuzzy and clustering methods have been introduced in this field since ancient times and have played important roles. Accordingly, in this paper, the Entropy weight-feature selection and ensemble learning is not used to diagnose cancer. Given the importance of breast cancer as well as the survival of patients with this cancer, this paper tries to use the Entropy weight-feature selection and ensemble learning, early breast cancer in people with premature ejaculation. When identifying and taking the necessary measures to eliminate it. Machine learning techniques make it possible to diagnose these types of diseases with low error rate, without the need for a specialist in this field, and in less time.

Therefore, in this paper, in section 2, the work done in the past is examined, in section 3, the proposed method and relevant details are described, finally in section 4, the results are obtained, and in section 5, the final conclusion and future suggestions are examined.

2- RELATED WORKS

Abu al-Allah Hassanian and his colleagues in 2014, a paper entitled "Using a diagonal division function and classifying multi-layered neural network networks" provided an MRI

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

method for diagnosing breast cancer. In this paper, they developed a combined approach that combines the benefits of fuzzy sets, derivative-based clustering, and classification of multi-layered proprioceptive neural networks (MLPNNs) with statistical-based extraction characteristics. The use of MRI imaging of breast cancer has been selected and the hybridization system has been applied to both benign or malignant outcomes to observe their ability and accuracy to classify breast cancer images. The introduced hybrid system begins with an algorithm based on fuzzy sets of type II to increase the contrast of the input images. This comes with an improved version of the classic ant-based clustering algorithm, called ant-based cluster maker, for identifying target objects through an optimization method that maintains the desired result during repetition. Then, more than twenty statistical-based features were extracted and normalized [6].

Bichen Zheng and colleagues in 2014, a paper titled diagnosis of breast cancer based on feature extraction using a combination of K-means and machine vector algorithm provided support. They concluded that the diagnosis of breast cancer was based on the characteristics of the tumor. Extracting features and options is critical to the quality of classifications created through data mining methods. A combination of K-means algorithm and support vector (K-SVM) has been developed to extract useful information and diagnose the tumor. The K-means algorithm is used to detect hidden patterns of benign and malignant tumors separately. The membership of each tumor is calculated according to these patterns and is considered as a new feature in the training model. A support vector machine (SVM) is then used to obtain a new classification for distinguishing input tumors [7].

Mary C. Poldon and colleagues in 2015, they presented a paper entitled "Weight gain" after the diagnosis of breast cancer and its mortality (systematic review and meta-analysis). They concluded that overweight and obesity were linked to breast cancer mortality. However, the association between weight gain and mortality after diagnosis is unclear. In a regular, meta-analysis of weight gain after breast cancer and specific cancer, they performed all mortality and recurrence outcomes. Cohort studies and clinical studies of weight change after diagnosis and mortality or recurrence of a specific disease in breast cancer were examined. Female participants with I-IIIC breast cancer were 18 years of age or older. Analysis of the fixed effects, the relationship between weight gain (≥5% body weight) and mortality in all cases; all experiments were bilateral [8].

ITags in 2015, he presented a paper entitled "The closest fuzzy-violent neighbor, along with an assessment of adaptability-based subsystems and sample selection for automatic diagnosis of breast cancer." "Breast cancer is one of the most common and deadly cancers for women," she said. Early diagnosis and treatment of breast cancer can increase patient outcomes. The development of high-precision classification models is a fundamental task in the field of medical information. Machine learning algorithms have been widely used to create strong and efficient classification models. In his paper, he presents a combined smart classification model

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

for diagnosing breast cancer. The proposed classification model consists of three steps: sample selection, feature selection, and classification. In sample selection, the fuzzy sampling method selection method is used based on the evaluation of the weak lost coefficient to eliminate inappropriate and incorrect cases. When selecting a feature, the compatibility-based feature selection method is used in relation to a re-ranking algorithm, given its efficiency in searching for possible titles in the search space. In the model classification phase, the closest fuzzy conjugation algorithm is used. Because this classification does not require the desired value for K neighbors and has the reliability value of richer classes, this method is used to perform the classification task [9].

Kriti Jiten Weir Warney and colleagues in 2015, they presented a paper entitled CAD PCA-PNN and PCA-SVM systems for classification of breast density. "Breast density is clinically important because there is a link between the risk of breast cancer and breast cancer," they predicted. In this paper, they compare the performance of two computer diagnostic systems (CAD) for classifying breast tissue density. This is done in the MIAS data set with 322 mammography images, including 106 greasy images and 216 dense images. ROIs are selected from one of the densest areas (eg, the center of each image, ignoring the stem muscle) of each mammogram. The total data set included 322 ROI (106 fat ROI and 216 dense ROI). The five statistical features of tissue are: mean, standard deviation, entropy, cortisone, and wrinkled images of the law's tissue energy, which are derived from the masks of Law 5, 7, and 9, are evaluated. Tissue vector calculations calculated from law masks of different lengths then analyze the main component (PCA) to reduce the dimensions of the desired space. SVM and PNN classification is used to perform the classification task. The promising results in the proposed CAD design show that it is useful in helping radiologists classify breast density [10].

Zirarata Kar and Di Dota Majomdar in 2016, they presented a paper entitled Early Diagnosis of breast cancer Using a New Approach to Mathematical Shape Theory and Fuzzy-Neural Classification System. They concluded: Breast cancer is one of the leading causes of death among women worldwide. For early detection and classification of abnormal breast growth in the breast or in the benign or malignant stage, they have used features based on mammography images. In their paper, they provide a comparative study of trait-based characteristics such as shape distance measurement and shape similarity of the area of interest from the mammography index to the neural-fuzzy classification system for pattern recognition [11].

R. Delhi Casillas Dewey and Dr. M. Indra Dewey in 2016, they presented a paper entitled Low Diagnosis Algorithm using wood classification to prevent premature breast cancer seizures. In this paper, they conclude that: Automatic diagnosis of breast cancer focuses on machine learning algorithms. The proposed approach has three stages of the process. In the first step, the data is grouped into a number of clusters using the first FAST algorithm for clustering. Due to the small size of the data set, the computational time is significantly reduced. In the second stage, the frequency of breast cancer data sets is identified using ODA. In the

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

third step, it determines whether the cancer is benign or malignant from pre-processed data using the J48 classification algorithm. The Wisconsin Breast Cancer Data (WBCD) and the Wisconsin diagnosis of breast cancer (WDBC) have been used to test the effectiveness of the proposed system [12].

Mehrabakhsh Nilshi and his colleagues in 2017, they presented a paper entitled A Knowledge-Based System for Classifying Breast Cancer Using Fuzzy Logical Method. They concluded: Breast cancer has become a common disease around the world. Expert systems are valuable tools that have been successful in diagnosing the disease. In this study, we create a new system based on a new knowledge-based system for classifying breast cancer using clustering, noise elimination, and classification techniques. (EM) is used as a clustering method for data clustering in similar groups. We then use CART classification and regression to produce fuzzy rules that are used to classify breast cancer disease in a knowledge-based system based on a fuzzy rule. To overcome the multi-column problem, we combine the main component analysis (PCA) in the proposed knowledge-based system. Experimental results on the diagnosis of viscous breast cancer and mammogram mass data collection show that the proposed methods significantly improve the accuracy of breast cancer prognosis. The proposed knowledge-based system can be used as a decision-making clinical support system to assist physicians in the practice of health care [13].

Fatemeh Ishaqi and her colleagues in 2017, they presented a paper entitled Classification Based on the Characteristics and diagnosis of breast cancer Using Fuzzy Inference System. They concluded that a model could be developed that could accurately predict the benign or malignant nature of the tumor. Because most biological systems are inherently fuzzy, the fuzzy inference system is suitable for classifying and diagnosing breast cancer based on the physical characteristics of the tumor. However, the function of the inferential fuzzy system is limited to a lack of understanding of the data on breast cancer. Therefore, in this paper, they present an advanced fuzzy inference system whose coefficients are optimized by statistical methods. Using a normalized set of cancer data, a fuzzy inference system is implemented to predict benign or malignant conditions. The results show that the optimized fuzzy inference system can more accurately control the problem of complex cancer prognosis [14].

Langrajorda et al., 2017, they presented a paper entitled Optimal Breast Cancer Classification using the Newton Gaussian algorithm. In this paper, they propose a new Gauss-Newton (GNRBA) rewriting algorithm for classifying breast cancer. It is used using a spartan dealership by selecting training samples. Until now, Nader has been known only in pattern recognition. The proposed method introduces a new Newton-based approach to obtaining the desired weight for training examples for classification. In addition, it compares Sparti to the common method of efficiency compared to the conventional l1-norm method. The effectiveness of GNRBA on the Wisconsin Breast Cancer Database (WBCD) and the Wisconsin Breast Cancer Database (WDBC) from the UCI machine learning repository have

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

been investigated. Various performance reports such as classification accuracy, sensitivity, properties, confusion matrices, a statistical test, and the area under the receptor factor (AUC) feature indicate the superiority of the proposed method over classical models. Experimental results suggest that the proposed GNRBA could be a viable alternative to classifying breast cancer for clinical experts [15].

Khan et al. (2019), in the paper, they examined the framework of e-health care services for the diagnosis and classification of breast cancer in cytological images. In this paper, a combination of machine learning and computational intelligence-based approaches was used as the Internet of Things Technology (IOMT) for early detection and classification of malignant breast cancer cells. The most innovative part of the proposed method is the use of algorithms (EA) to select the optimal features that reduce computational complexity and accelerate the classification process in cloud-based e-care services. The performance of the proposed method has been confirmed by experiments on real data sets that provide 98.0% accuracy in the diagnosis and classification of malignant cells in cytological images [16].

Osmanovic et al. (2019), the paper looked at machine learning methods for classifying breast cancer. In this paper, a different diagnostic system is designed and tested to diagnose breast cancer patients based on samples that characterize the cell nuclei in the digital image from fine needle aspiration (FNA). The results showed that a hidden single-layer neural network with 20 neurons had the highest classification accuracy (98.9% and 99% accuracy in the test). The accuracy of multilayer architectures was significantly lower, ranging from 86.3 to 74.9%, with an average value of 81.37%. A carefully developed advanced system can be used in the laboratory as a promising method for early diagnosis of breast cancer in the future [17].

Asharia et al. (2020), In the paper, they examined the deep network for classifying breast cancer and increasing loss of function (ELF). The aim of this study was to control diagnostic error by increasing image accuracy and reducing processing time using several algorithms such as deep learning, K-means and clustering. Histopathological images were obtained from five data sets and pre-processed using normalization and linear conversion filter. The present study showed that the deep learning algorithm increased the accuracy of breast cancer classification by 97% compared to the advanced model, which accounted for 95%, and the change time was reduced from 30 to 40%. It has also increased system performance by improving clustering using K-means for nonlinear histopathological image conversion [18].

Wang et al. (2020), in the classification paper, images of breast cancer with deep stimulation characteristics were examined. In this paper, a hybrid structure that includes a Dual Transfer Learning Dual Learning (D2TL) and Interactive Learning Machine (ICELM) is presented based on the ability to extract and demonstrate ability. First, high-level features are extracted using two-step deep transfer learning. Then, high-level attribute sets are commonly used as regulatory terms to further improve classification performance in cross-functional learning

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

devices. The results show that the proposed method has had a significant performance in classification accuracy (96.67%, 96.96%, 98.18%). From the results of the experiment, the proposed method promises to provide an effective tool for classifying breast cancer in clinical settings [19].

Sojata et al. (2020), in one study, they screened and identified primary breast cancer using tissue-based ANFIS classification. The aim of this study is to develop strong diagnostic schemes for early detection of breast cancer. This is a non-invasive procedure called adaptive fuzzy neural structure (ANFIS) for the diagnosis of microscopy (cancerous lesions) in the mammary glands. Clinical images The MIAS database is used to test and teach the proposed algorithm. ANFIS-based classification is an effective way to diagnose breast cancer. The performance of the proposed method (tissue characteristics with ANFIS) is compared with the outputs obtained from the K-means clustering algorithm, wavelet transform and artificial neural network classification using the mean absolute deviation as a feature to detect early stages [20].

Siraham et al. (2020), In one study, they classified breast cancer thermographs with entropy features with neural networks. In this study, the classification of chest thermographs using a possible neural network (PNN) with statistical quadrilateral, mean, standard deviation, skewness and elongation and two features of entropy, Shannon's entropy and closed entropy were examined. The CLAHE histogram equation algorithm with uniform and rail distribution was considered to enhance the contrast of breast thermal images. The simulation study shows that CLAHE-RD with entropy characteristics of the wave confirms the presence of symmetry in the thermal images of the right and left breasts. The overall classification accuracy of 92.5% was obtained using several proposed versions with PNN classification. It thus confirms the appropriate proposed method as a screening tool for the diagnosis of asymmetry as well as the classification of breast thermographs [21].

Gonzalez et al. (2020), in the paper, they presented a new model of artificial immune system for classifying communication with competitive performance for the diagnosis of breast cancer. The Wilcoxon test was used to identify statistically significant differences between the proposed method and other classification algorithms based on the bio-inspired model. These statistical tests show better performance shown by the proposed model with better than other immune system-based algorithms. The proposed model is competitive with other known classification models. In addition, this model has a low computational cost. The success of this model in classification work shows that collective intelligence is useful for this type of problem [22].

In a paper, Mushtaq et al. (2020) examined breast cancer data sets using the k-class closest neighbor. The function of KNN depends on the number of neighboring elements known as K values. This study includes exploration of KNN performance using different distance functions and K values to find an effective KNN. Countless evaluation criteria, such as accuracy,

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

characteristic curve of the receiver (ROC) with area below the curve (AUC), and sensitivity, etc., were used to evaluate the implemented techniques. The results showed that the technique of selecting the feature using the distance functions was the most accurate for both data sets. Proper selection of k value and a distance function in KNN, a feature selection for the Cancer Data Collection, shows the highest accuracy compared to existing models [23].

Table 1: Summary of proposed methods

Authors	Year	Method	Result
Hassanian et al	2014	The use of diagonal-based segmentation function and classification of multilayer neural network networks, MRI method for the diagnosis of breast cancer	The introduced hybrid system begins with an algorithm based on fuzzy sets of type II to increase the contrast of the input images.
Zheng et al	2014	Diagnosis of breast cancer based on feature extraction using a combination of K-means and support vector machine algorithm	Extracting features and options is critical to the quality of classifications created through data mining methods.
Poldon et al	2015	Systematic and meta-analysis of weight gain after diagnosis of breast cancer and its mortality	Analysis of the fixed effects, the relationship between weight gain (≥5% body weight) and mortality of all cases; all experiments were bilateral.
I tag them	2015	The closest fuzzy-rough neighbor, along with a compatibility-based subsystem assessment and sample selection for automatic diagnosis of breast cancer	This classification does not require the desired value for K neighbors and has the reliability value of richer classes. This method is used to perform the classification task.
Jiten et al	2015	Provides CAD PCA-PNN and PCA-SVM systems for chest compaction classification	The promising results in the proposed CAD design show that it is useful for helping radiologists classify breast density.

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

Zirarata Kar et al	2016	Early diagnosis of breast cancer using a new approach to mathematical shape theory and fuzzy-neural classification system	Features based on mammography images for early detection and classification of abnormal growth of lumps in the breast or in the benign or malignant stage
Casillas Dewey et al	2016	Low diagnostic algorithm using wood classification to prevent premature breast cancer seizures	Automatic diagnosis of breast cancer focuses on machine learning algorithms.
Nilshi et al	2017	A knowledge-based system for classifying breast cancer using fuzzy logic	The proposed methods significantly improve the accuracy of breast cancer prognosis.
Ishaqi et al	2017	Feature-based classification and diagnosis of breast cancer using fuzzy inference system	The results show that the optimized fuzzy inference system can more accurately control the problem of complex cancer prognosis.
Langrajorda et al	2017	Optimal breast cancer classification using Newton's Gaussian algorithm	Experimental results suggest that the proposed GNRBA could be a viable alternative to classifying breast cancer for clinical experts.
Khan et al	2019	Evaluate the framework of e-health care services for the diagnosis and classification of breast cancer in cytological images	The performance of the proposed method has been confirmed by experiments on real data sets that provide 98.0% accuracy in the diagnosis and classification of malignant cells in cytological images.
Osmanovich et al	2019	Investigating machine learning methods for classifying breast cancer	The results showed that a hidden single-layer neural network of backup processing with 20 neurons had the highest classification accuracy.

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

Asharia et al	2020	An in-depth study of breast cancer classification and increased loss function (ELF(	The present study showed that the deep learning algorithm increased the accuracy of breast cancer classification by 97% compared to the advanced model, which accounted for 95%.
Wang et al	2020	Investigating the classification of breast cancer images with deep stimulation features	The results show that the proposed method has had a significant performance in classification accuracy.
Sojata et al	2020	Screening and early detection of breast cancer using tissue-based ANFIS classification	Better performance of the proposed method using neural network-based classification method using mean deviation
Siraham et al	2020	Classification of breast cancer thermographs with entropy features with neural network	Properly recommended as a screening tool to diagnose asymmetry as well as classify breast thermographs.
Gonzalez et al	2020	Provide a new artificial immune system model for classifying communication with competitive performance to diagnose breast cancer	The results show that the proposed method performs better than other algorithms.
Mushtaq et al	2020	Investigate the collection of breast cancer data using the nearest k classification	Choosing the right amount of k and a distance function in KNN, the feature selection for the Cancer Data Collection, shows the highest accuracy compared to existing models.

3- THE PROPOSED METHOD

In this section, we will describe and present the proposed method and then the steps in the flowchart will be fully explained along with the details.

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

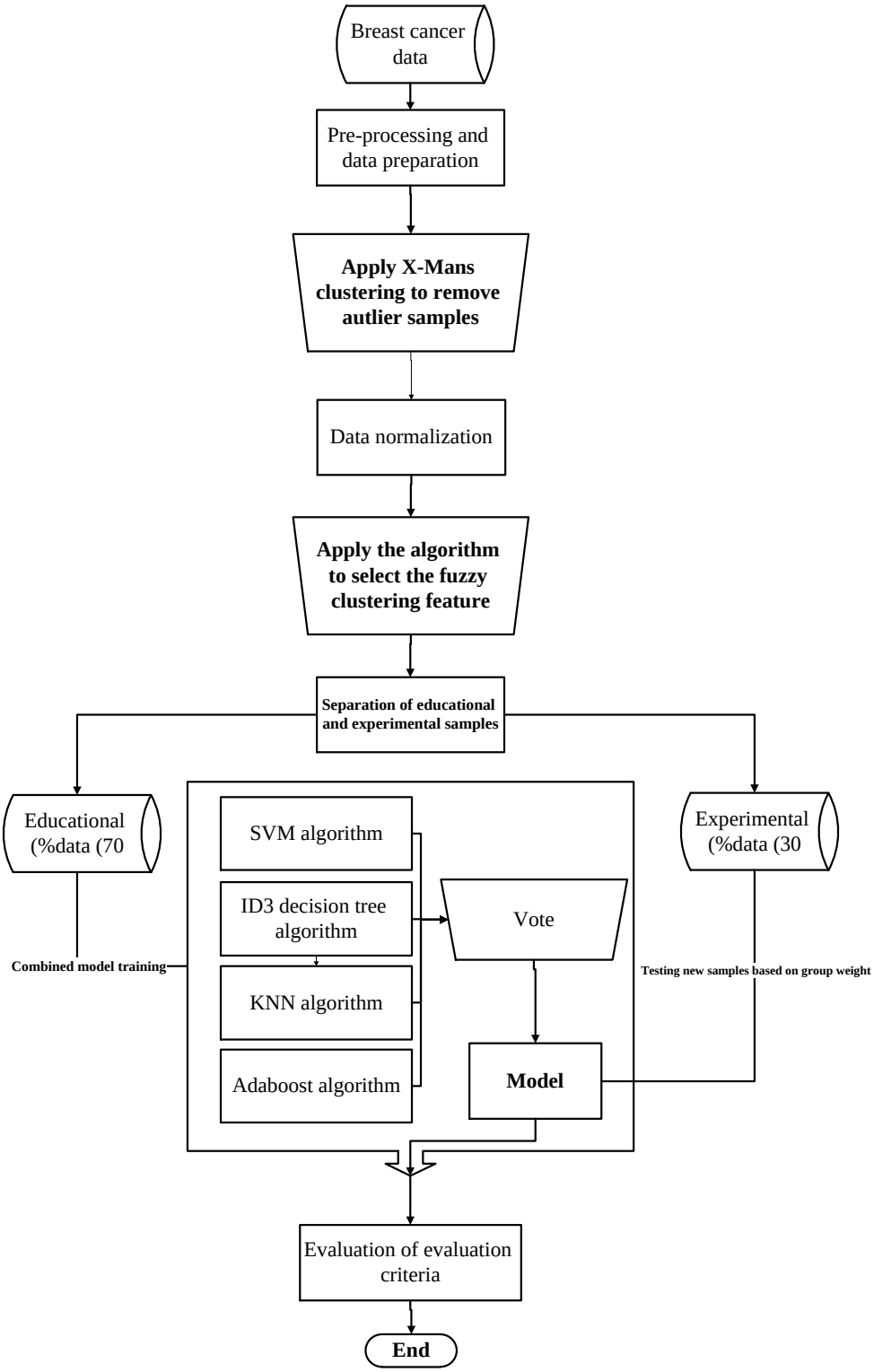

Figure 1: Flowchart of the proposed method

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

As can be seen from the flowchart figure (1), first the breast cancer dataset received from the popular SEER system is entered into the proposed system. Given that this data is not standardized and coherent, it is necessary to make a pre-processing on this data and make it coherent. Therefore, at this stage, using data cleansing techniques and deleting discarded values, unused data are removed from the database and the data is prepared for further execution. Clustering is then performed on all breast cancer data in women using the X-Means algorithm. Using the clustering technique, we can identify the samples that are discarded and by removing them, we assign the final labels to each sample. The algorithm we used for clustering is the X-Means algorithm. We then look at which clusters have the least impact on outcomes and survival rates. We remove these clusters from the list of data. In the next step, the data are normalized to produce more accurate models and increase the accuracy of cancer diagnosis.

The standardized and coherent data are then entered into the algorithm core of the selection-weighted-entropy fuzzy-based fuzzy clustering feature. This algorithm executes the feature selection process on coherent data and selects the features that have the most impact on the output. The next step is to separate experimental and training samples to classify and diagnose the survival rate of breast cancer, which in this paper is 70% to 30%. This means that 70% of the main data on which the algorithm is selected using the fuzzy cluster-based fuzzy clustering feature-weight will be considered 30% as testing samples. In the next step, the instructional examples are applied to the combined learning system, and each machine learning algorithm, including SVM, Adaboost, KNN, ID3, produces models separately. The experimental data are then applied to the produced models and finally the cancer diagnosis process takes place. Finally, the evaluation criteria are applied to the results and the accuracy, error, accuracy and recall are calculated.

3-1- Pre-processing data

After the data is entered into the proposed system, the data is pre-processed and the discarded and unused samples are removed. The data, which then became coherent after pre-processing, is then converted to an acceptable format for simulation tools. At this stage, the data is usually converted to Excel and coherent format. Various methods have been proposed to apply pre-processing to the data, which are:

- Data Cleanup
- Data Aggregation
- Data Transitions
- Data Reduction

Given the needs of our issue in this paper, we have only used the data cleansing method. The proposed strategy is to analyze the data and identify if the row or column had empty or

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

unused values. We then examine the values before and after the sample that has a blank or unused value and calculate their mean. Finally, we will replace the empty value with the average obtained. This eliminates waste samples and produces more consistent data.

3-2- Data preparation

We also need to prepare the data after the discarded samples are lost. To do this, we convert the pre-processing data into an acceptable format for simulation tools. The default data format is Excel. So that we have analyzed the data superficially, it must be behavioral and analyzed.

3-3- Remove outlier data using the x-means clustering algorithm

Clustering is one of the methodological methods used specifically in distinguishing distinct clusters within data. It should be noted that the clustering method is used to remove discarded data. Without data mining, the analyst must visualize the data to distinguish between distinct clusters and identify the index relationships in each cluster. In this case, the risk of ignoring important categories of data is very high. Using data mining, it is possible to allow the data itself to display groups between them. This is one of the black box methods of data mining algorithms that is difficult to understand. Machine learning algorithms can show different clusters in a data set.

In this paper, X-Means clustering technique is used to create coherence and determination for each sample due to the fact that the data are unattended and do not have a suitable category. In this case, we applied the X-Means clustering algorithm to a file containing four cancers. This file contained approximately eight thousand data with twelve features. Once the relevant cluster has been determined for each sample, we remove the inappropriate data and find the hidden patterns in the data to improve the results of the diagnosis.

The following algorithm shows the X-Means code.

X-means Algorithm Steps:

(1) Initialize K = Kmin.

(2) Run K-means algorithm.

(3) FOR k = 1,. . . ,K: Replace each centroid μk by two centroids μ(1) and μ(2).

Algorithm 1. X-Means clustering algorithm

In general, this algorithm first calculates the error rate of clustering according to the data dimensions and determines which number of clusters has the lowest error rate. After determining the number of relevant ks that are optimal, it executes the K-means algorithm and performs clustering. In fact, this algorithm is an improved state of the K-Means algorithm.

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

Therefore, we use the X-Means clustering algorithm to remove discarded data to increase the accuracy of the sample classification. Then the data is normalized.

3-4- Data normalization

In the preprocessing step, in order to get better results, we normalize the values of each feature of the used dataset between 0 and 1, then we randomly move the general data matrix rows so that the data are arranged from the initial state of the sum. Gathered, get out. In other words, all the data is written in the form of a matrix and the normalization of the rows of the matrix is done. Normalization is due to higher accuracy. The following equation is used to normalize the values of each data set.

$$\text{Normalize}(x) = \frac{(\text{x} - \text{X}_{\text{min}})}{(\text{X}_{\text{max}} - \text{X}_{\text{min}})} \quad (1)$$

Xmax and Xmin are the maximum and minimum values in the X feature range. After normalizing the data, the values of all the attributes are in the range [0.1]. After normalizing the relevant data, a pre-processing is applied to the existing samples and the data that have discarded and unused values are deleted. In addition, samples from which no activity or function has been recorded will be cleared from the original database and the result will be sent to the next section. Therefore, after pre-processing and preparation of the initial data and applying the normalization technique on the data, the feature selection process is performed using Entropy weight-feature selection and ensemble learning, which is described in the next section.

3-5- Entropy weight-feature selection

Once the data has been pre-processed and normalized, it is time to select the features that are effective in determining and diagnosing the survival rate of breast cancer. This process is done with the help of fuzzy clustering and weight-feature entropy. We want to measure the amount of information that occurs in an event. This is a function of p if a p is likely to occur. In fact, entropy is an event that measures the amount of information that does not depend on the event but on its probability. Some conceptual considerations regarding information suggest that this function is declining in terms of p and for a definite event its value is zero. With (E (p showing), we want the continuous E function to apply to the following conventional principles:

$$\text{E}(p) \downarrow$$

$$\boldsymbol{E(pq) = E(p) + E(q)} \quad (2)$$

$$\boldsymbol{E(1)}=0$$

The answer to the function that applies in the above conditions is uniquely equal

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

$$E(p) = clnp \quad (3)$$

According to the contract, c=-1 is taken over.

If the random variable X assumes the values x2, x1, etc., then $A_i = X^{-1}(x_i)$ will be a separation of the sample space ‹, in which case the entropy X with the symbol E (X) According to the mathematical hope of entropy, the events of A_i are defined:

$$E(X) = -\sum p_i lnp_i \quad (4)$$

where in $P_i = P(A_i)$. Since entropy depends only on probabilities and not on X values for ease and some other uses, we may also display entropy with the following symbol:

$E(p_1, p_2, \ldots)$ (5)

It is easy to see that the entropy defined above has the following characteristics:

1. *E* is a function of the probabilities *p1*, *Pn*.
2. *E* is a function of each of its components.
3. *E* is symmetrical with respect to each of its components, meaning that the value of *E* does not change as the components change.
4. If a product with zero probability is added to the set of outputs, *E* will not change.
5. If there is no uncertainty, *E* must be minimized.
6. If the uncertainty is maximum, *E* must also be maximum.
7. Increase the maximum value of *E* by n.
8. For two independent probability distributions such as $p=(p_1, \ldots, p_n)$ and $q=(q_1, \ldots, q_m)$ the uncertainty of these two systems must be equal to the sum of the uncertainties of each.
9. If the two systems are not independent, the uncertainty of the two systems is calculated as follows: The uncertainty of the two systems is equal to the sum of the uncertainty of one of the systems and the mathematical hope of the condition of the uncertainty of the second system when the first system is realized. Another $E(p \cup q) = H(p) + H(q|p)$

The command $E(p_1.p_2.\ldots) = -\sum p_i lnp_i$ is a command that was proposed by Shannon and proved to be the only command that applies to the above nine attributes. Many scientists have tried to use some of the above features as the basis for other entropy commands, some of the most common of which are listed below:

Shannon (1994) $E(p) = -k\sum p_i lnp_i$

Renny's entropy (1961) $E_\alpha(p) = \frac{1}{1-\alpha} ln \sum p_i^\alpha$

Kapoor (1967 and 1968) $E_{\alpha,\beta} = \frac{1}{1-\alpha} ln \sum \frac{\sum p_i^{\alpha+\beta-1}}{\sum p_i^\beta}$

Hawarala and Charwat $E^\alpha(\mathrm{p}) = \frac{\sum \mathrm{p_i^\alpha - 1}}{21-\alpha-1} = \frac{1}{1-\alpha}\left[\mathrm{e}^{(1-\alpha)\mathrm{E_{d(p)}}} - 1\right]$

This is a non-collective size. $H^\alpha(\mathrm{p} \cup \mathrm{q}) = H^\alpha(\mathrm{p}) + \mathrm{H}^\alpha(\mathrm{q}) + (1-\alpha)\mathrm{H}^\alpha(\mathrm{p})\mathrm{H}^\alpha(\mathrm{q})$

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

And other non-collective sizes defined by other individuals (about 12 types of non-collective sizes).

It is easy to show that:

$$\lim_{a\to 1} E_a(p) = E(p)$$
$$\lim_{a\to 1} E_{a.1}(p) = \lim_{a\to 1} E_a(p) = E(p)$$
$$\lim_{a\to \infty} E_{a.1}(p) = -\ln \max\{p_1.p_2.p_3.\ldots.p_n\} \quad (6)$$

The beauty of what has been done about discrete variables cannot be determined by the entropy size for continuous variables. One reason for this is that the number of component states is usually finite in applied systems, so what was done with discrete variables was natural and tangible. Discrete as equation (7):

$$E(f) = -\int_{-\infty}^{\infty} -\infty f(x)\ln f(x)dx \quad (7)$$

some ideas of disconnecting from disconnected mode can be a motivator for high definition.

Thus, entropy makes it possible to derive the probability of the effect of one or more properties or variables on output based on probability. Therefore, we will use entropy properties as the rules for selecting properties, and we will cluster each sample for one to several features based on the probability. All clusters have proportional values that are evaluated by a proportional function for optimization. In order to implement our proposed approach to select the optimal features in this paper, the following proportionality function is used in the clustering process in order to find the effect and ultimately the accuracy of classification of each feature against the cancer survival rate.

$$fitness_i = w_A \times acc_i + w_F \times \left[1 - \frac{\left(\sum_{j=1}^{n_F} f_i\right)}{n_F}\right] \quad (8)$$

w_A shows the weight accuracy of fuzzy clustering to select a feature. acc_i The accuracy of the clustering algorithm, w_F weight is selected for the number of features, f_i shows the value of the attribute mask, the value of "1" indicates that the attribute j is selected, and the value of "0" indicates that the attribute j is not selected. Is. n_F shows the total number of properties. The following equation is used to calculate acc_i, which shows the accuracy of fuzzy clustering.

$$acc = \frac{cc}{cc + uc} \times 100\% \quad (9)$$

cc indicates the number of correctly categorized samples and *uc* indicates the number of incorrectly classified samples. The following are steps in selecting outstanding features by the Entropy weight-feature selection Algorithm. The logic of the alignment function of the proposed feature selection algorithm is described below.

(p_5،p_4،p_{32}،p،p) = attributes

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

$(D_5‘D_4‘D_3‘D_2‘D_1)$ = data

Data selection of individual features and calculation of each accuracy.

(p_5) (p_4) (p_3) (p_2) (p_1)

Select a set of duplicate attributes and calculate the accuracy of each.

,…, $(p_4, p_5))$ $p_3‘p_2)$ $(p_5‘p_1)$ $(p_4‘p_1)$ $(p_3‘p_1)$ $(p_2‘p_1($

As well as selecting multiple features to the total number of features.

$(p_5‘p_4‘p_3‘_2p‘p_1)$

Finally, compare the accuracy of each and select the feature of the set that has the highest accuracy. According to the calculation of accuracy in each step and the calculation of the fit of each repetition according to the accuracy of the classification, the process of selecting the superior feature is performed. Therefore, after selecting the features that have the greatest impact on the diagnosis of cancer, it is time to process the separation of training and experimental data and finally classify the combined samples.

3-6- Training and testing samples

One of the important phases that must be applied to teach machine learning algorithms in the proposed system is the separation of training samples to teach the model of the algorithms in the proposed system and samples as testing samples to test the proposed method.

Sampling of the desired data is one of the steps of data mining that has been considered in this paper. There are several methods of sampling, the three most important of which are:

- Random Sampling
- Stratified Sampling
- Balance Sampling

Random sampling is one of the simplest sampling methods that works randomly and segregates samples from the main data as training and experimental data. One of the disadvantages of this method is that it is not possible to sample any particular category and ultimately leads to a decrease in the accuracy of data classification and validation of the proposed method. Classified sampling is also one of the improved methods of random sampling. This method performs the sampling process based on probability and also selects the samples as a percentage. This sampling method is also difficult and may not select probability-based samples. Balanced sampling is one of the methods that selects the required samples in a balanced way from the existing categories and classes. This method does not have the problem of the previous methods and finally selects the balanced data and samples between the existing samples. The method used to separate training data and experimental data is the Balancing method. Therefore, 70% of the samples are separated as training samples and 30% as testing samples.

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

3-7- Diagnosis of breast cancer using ensemble learning

Once an optimal dataset with a set of salient features has been formed, it is necessary to enter the training data one by one into the Adaboost, SVM, ID3, and KNN algorithms. After entering the training data, each of these algorithms produces its own models separately. After the relevant models have received the necessary training, they produce a set of rules so that in the next steps, the test samples are entered into the model and the survival rate in breast cancer is determined.

In the proposed method, each algorithm sends its response to the Vote system independently. At the core of the Vote system is a survey of the responses of other algorithms and considers the number of more responses. Finally, the answer that has more votes from SVMM, Adaboost, KNN, ID3 algorithms is considered as the main answer. Once it has been determined whether the sample has cancer, the final dataset is examined to see if there is another sample that is identical to the current sample. If there is a sample, it is clear that it is a cancer and the survival rate of the cancer is much lower for people, because having more than one type of cancer in one person reduces the survival of the person. If a person does not have cancer, the coefficient has no effect on it and the survival rate of the person increases. In this strategy, for each new patient who enters the model, the presence of cancer is first examined.

4- EXPERIMENTAL RESULTS

First, we will compare the proposed method with other methods in this section, first lay out the test settings, then set out the benchmarks, and then analyze the experiments.

4-1- Experimental setup

The proposed method in this paper is implemented using Matlab simulator version 2015a. Also, the operating system is Windows 10, the operating system type is also the 32-bit operating system, 4GB of RAM used - 3.06GB usable, Intel processor - Number of cores 7 (Core ™) i7 CPU) - Q 720 @ 1.60GHz is 1.60 GHz.

4-2- Dataset

The first step in the proposed method is to enter the dataset into the proposed system. Therefore, in this paper, we will use the SEER dataset. We use the SEER database, a global cancer institute that has studied American data. The National Cancer Institute's Monitoring, Epidemiology, and Final Outcomes (SEER) program provides statistical information on cancer to reduce the burden of cancer among the U.S. population. SEER collects cancer data from various locations and sources across the United States. Data collection began in 1973 with a limited amount and has continued to expand to more areas. The collection of SEER data from

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

cancer patients and their survival began on January 1, 1973, in the states of Connecticut, Iowa, New Mexico, Utah, Hawaii, and the urban areas of Detroit, San Francisco, and Auckland. In 1974-1975, the Atlanta metropolitan area and 13 counties in the Seattle-Puget Sound region were added to the data, and in 1978, 10 black and white rural counties in Georgia. In addition, American Indians living in Arizona were added in 1980. Three other geographical areas were added to the SEER program before 1990: New Orleans, Louisiana (between 1974 and 1977 and again in 2001), New Jersey (between 1979 and 1989, and again in 2001) and Puerto Rico (1973). Until 1989). Also with the Cancer Registration Budget from the National Cancer Institute and with the technical assistance of SEER, data collection on cancer cases among indigenous peoples of Alaska living in Alaska began.

In 1992, the SEER program to increase the coverage of the minority population, especially the Spanish, expanded with the addition of Los Angeles and four counties in the San Jose-Monteri area south of San Francisco. In 2001, the program also covered SEER, Kentucky and the remaining cities in California (Greater California). In addition, New Jersey and Louisiana cancer data were once again added to the data. In 2010, coverage of the SEER program expanded to include all parts of the state of Georgia [24]. The SEER database collects and publishes data on various cancers. The study population covered about 28 percent of the U.S. population. SEER coverage covers 26 percent of African Americans, 38 percent of Hispanics, 44 percent of Native Americans and Alaska Indigenous peoples, 50 percent of Asians, and 67 percent of Hawaii, the Pacific Islands. Based on the definition of coherence due to the occurrence of diseases, the samples were limited to those patients who had exactly two cancers and their cancer was diagnosed within 1 year. Thus, the number of male and female specimens was 3664 and 14243, each of which has 149 variables (characteristics). The number of samples used in this paper is about 20,000 patients, both healthy and infected. Therefore, in this paper, this popular database is used in the same conditions as the comparative research.

4-3- Performance metrics

In this paper, these two tools are used for data analysis and simulation. Also, the criteria used in this paper to evaluate the proposed method by other methods are:

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Equation (10) is used to check the accuracy of the proposed method. The TP (True Positive) parameter indicates the number of samples that have been correctly identified as intrusive. The FP (False Positive) parameter also indicates the number of samples that have been mistakenly detected.

$$ReCall = \frac{TP}{TP + FN} \tag{11}$$

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

Equation (11) indicates the call rate of the proposed method where the FN (False Negative) parameter indicates the number of samples that have been diagnosed as normal.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

Equation (12) is used to calculate accuracy where the only parameter not described is TN (True Negative), which represents the number of samples that have been correctly identified as normal. Finally, the error rate of the proposed method is calculated based on (13).

Also, equation (14) is used to calculate the Mean Absolute Error (MAE) metric.

$$Error = 100 - (TP + TN)/(TP + TN + FP + FN) \quad (13)$$

$$MAE = Mean(Error^2) \quad (14)$$

The Root Mean Square Errors (RMSE) is calculated by equation (15).

$$RMSE = \sqrt{Error} \quad (15)$$

By simulating the proposed model in the IDS, the results are obtained based on the evaluation criteria presented.

4-4- Analyze of results

In this section, we evaluate the results obtained from the simulation of the proposed method. We define three scenarios for evaluating the results. In this section, we analysis of the results of the proposed method. Table (3) shown results of base machine learning algorithm and ensemble learning system.

Table 3. Results of base machine learning algorithm and ensemble learning system

	KNN	**SVM**	**Adaboost**	**ID3**	**Entropy-Ensemble**
TP	3167	3176	3336	4733	4733
TN	200	0	1607	628	56
FP	1566	1557	1397	0	0
FN	13432	13632	12025	23004	13576
Accuracy	90.38%	91.52%	83.66%	96.58%	**99.69%**
Error	9.62%	8.48%	16.34%	3.42%	**0.31%**
Recall	82.73%	83.57%	79.10%	97.69%	**99.79%**
Precision	91.81%	94.88%	73.99%	94.16%	**99.41%**

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

MAE	0.301	0.152	0.160	0.035	**0.033**
RMSE	0.342	0.276	0.296	0.183	**0.092**

As can be seen from the Table (3), the accuracy of diagnosis of breast cancer using KNN algorithm is 90.38%. Breast cancer diagnosis error rate is 9.62%. The recall and precision metric of the KNN algorithm for the diagnosis of breast cancer in infected people is 82.73% and 91.81%, respectively. The MAE and RMSE of KNN algorithm are 0.301 and 0.342, respectively. The accuracy of diagnosis of breast cancer using SVM algorithm is 91.52%. Breast cancer diagnosis error rate is 8.48%. The recall and precision metric of the SVM algorithm for the diagnosis of breast cancer in infected people is 83.57% and 94.88%, respectively. The MAE and RMSE of SVM algorithm are 0.152 and 0.276, respectively. The accuracy of diagnosis of breast cancer using Adaboost algorithm is 83.66%. Breast cancer diagnosis error rate is 16.34%. The recall and precision metric of the Adaboost algorithm for the diagnosis of breast cancer in infected people is 79.10% and 73.99%, respectively. The MAE and RMSE of Adaboost algorithm are 0.160 and 0.296, respectively. The accuracy of diagnosis of breast cancer using ID3 decision tree algorithm is 96.58%. Breast cancer diagnosis error rate is 3.42%. The recall and precision metric of the ID3 decision tree algorithm for the diagnosis of breast cancer in infected people is 97.69% and 94.16%, respectively. The MAE and RMSE of ID3 decision tree algorithm are 0.035 and 0.183, respectively. The accuracy of diagnosis of breast cancer using the proposed method that called Entropy-Ensemble is 99.69%. Breast cancer diagnosis error rate is 0.31%. The recall and precision metric of the Entropy-Ensemble algorithm for the diagnosis of breast cancer in infected people is 99.79% and 99.41%, respectively. The MAE and RMSE of Entropy-Ensemble algorithm are 0.033 and 0.092, respectively.

As can be seen from the Table (3), the accuracy of the Entropy-Ensemble algorithm compared to ID3, Adaboost, SVM and KNN decision tree algorithms is about 3.11%, 16.03%, 8.44% and 9.31%, respectively. The recall of the Entropy-Ensemble algorithm compared to ID3, Adaboost, SVM and KNN decision tree algorithms is 2.1%, 20.69%, 16.04% and 17.06%, respectively. The precision of the Entropy-Ensemble algorithm compared to ID3, Adaboost, SVM and KNN decision tree algorithms is about 5.25%, 25.42%, 4.53% and 7.6%, respectively. The MAE rate of Entropy-Ensemble algorithm compared to ID3, Adaboost, SVM and KNN decision tree algorithms is about 0.002%, 0.127%, 4.53% and 0.119%, respectively. The RMSE of the Entropy-Ensemble algorithm compared to ID3, Adaboost, SVM and KNN decision tree algorithms is about 0.091%, 0.177%, 0.243% and 0.25%, respectively

Due to the simulation of the proposed method in this paper with other methods in other researches in the same conditions and data, in this section, the results obtained from the

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

proposed method are evaluated and compared with the paper [26]. The Fig. 2. compares the accuracy of the proposed method with other methods.

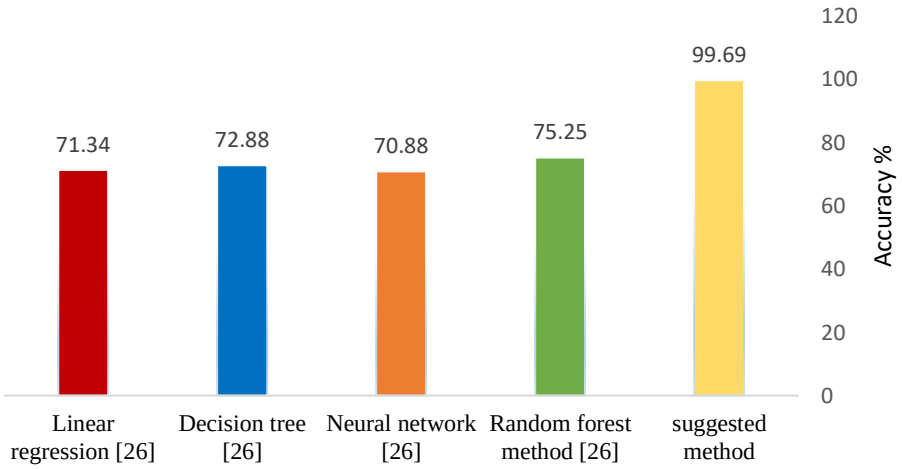

Figure 2. Comparison of the accuracy of diagnosis of breast cancer in the proposed method compared to other methods

The accuracy of the Entropy-Ensemble algorithm for diagnosing breast cancer compared to Linear regression, decision tree, neural network and random forest that discussed in [26] is about 24.44%, 28.81%, 26.8% and 28.35% respectively. The following figure shows the accuracy of the diagnosis of breast cancer in the proposed method compared to other methods.

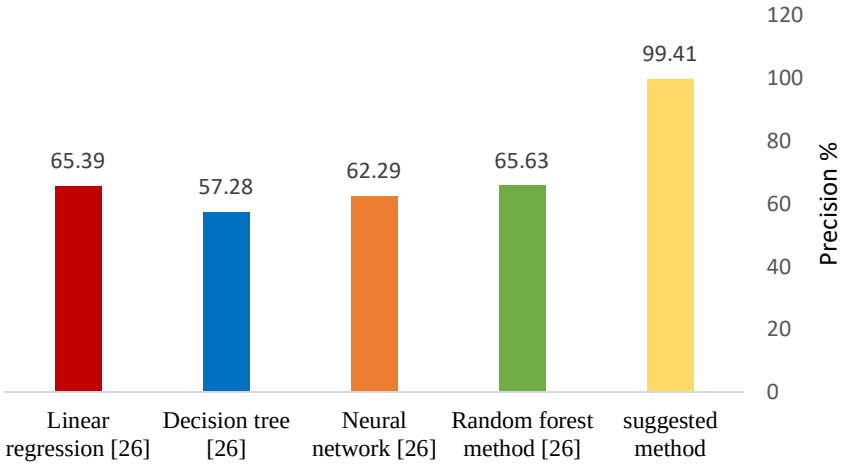

Figure 3. Comparison of the precision of diagnosis of breast cancer in the proposed method compared to other methods

As can be seen in the Fig. 3, the precision of the Entropy-Ensemble algorithm for diagnosing breast cancer compared to Linear regression, decision tree, neural network and random forest that discussed in [26] is about 24.44%, 28.81%, 26.8% and 28.35%, respectively. The Fig. 4.

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

shows a comparison of the recall for diagnosis of breast cancer in the proposed method compared to other methods.

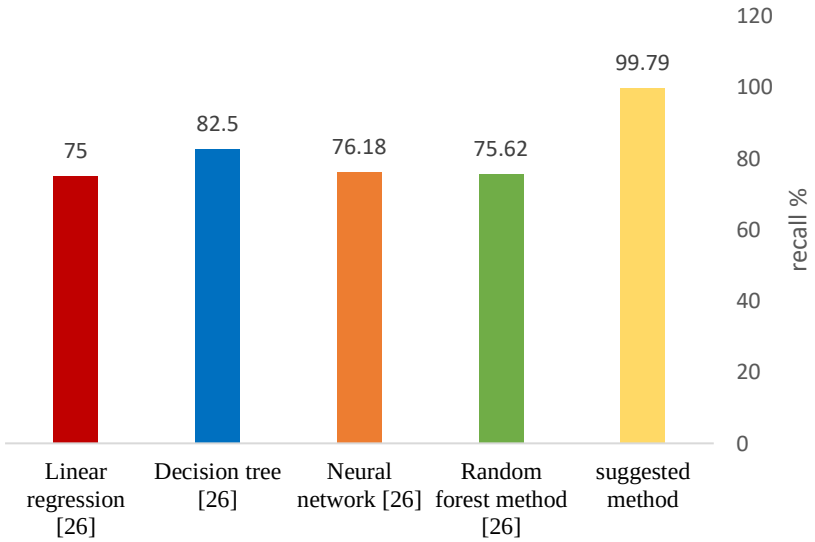

Figure 4. Comparison of breast cancer detection recall in the proposed method compared to other methods

As can be seen in the Fig. 4, the recall of the Entropy-Ensemble algorithm for diagnosing breast cancer compared to Linear regression decision tree, neural network and random forest that discussed in [26] is about 24.79%, 17.29%, 23.61% and 24.17%, respectively. The Fig. 5. shows a comparison of the error rate for diagnosis of breast cancer in the proposed method compared to other methods.

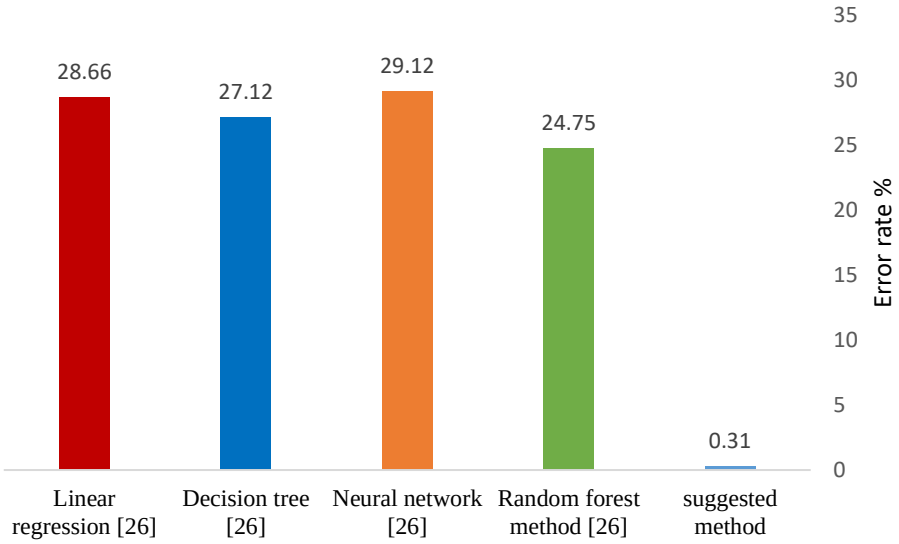

Figure 26: Comparison of diagnosis of breast cancer error in the proposed method compared to other methods

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

The error rate of the Entropy-Ensemble algorithm for diagnosing breast cancer compared to Linear regression, decision tree, neural network and random forest that discussed in [26] is about 24.44%, 28.81%, 26.8% and 28.35% respectively.

5- Conclusion

Based on the simulation, it was observed that the accuracy, precision, recall and error detection of breast cancer diagnosis in the proposed method compared to other methods has improved significantly. The Table (3) are the final results of the improvement of the proposed method compared to the methods presented in the paper [26].

Table 3. Results of the improvement of predictive accuracy in the proposed method compared to other methods

Entropy-Ensemble	Random forest [26]	Neural network [26]	Decision tree [26]	Linear regression [26]
99.69	75.25	70.88	72.88	71.34

According to the table above, it can be seen that the rate of improvement of predictive accuracy in the proposed method has significantly improved compared to other methods.

Reference

[1]. Chaurasia, V., & Pal, S. (2014). Data mining techniques: To predict and resolve breast cancer survivability. *International Journal of Computer Science and Mobile Computing*, *3*(1), 10-22.

[2]. R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2012," *CA: a cancer journal for clinicians*, vol. 62, pp. 10-29, 2012.

[3]. J. G. Elmore, C. K. Wells, C. H. Lee, D. H. Howard, and A. R. Feinstein, "Variability in radiologists' interpretations of mammograms," *New England Journal of Medicine*, vol. 331, pp. 1493-1499, 1994.

[4]. Fehrenbacher, L., Capra, A. M., Quesenberry Jr, C. P., Fulton, R., Shiraz, P., & Habel, L. A. (2014). Distant invasive breast cancer recurrence risk in human epidermal growth factor receptor 2–positive T1a and T1b node-negative localized breast cancer diagnosed from 2000 to 2006: A cohort from an integrated health care delivery system. *Journal of Clinical Oncology*, *32*(20), 2151-2158.

[5]. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. J. Artificial Intelligence in Medicine 2010; 34: 113-27.

[6]. Aboul EllaHassanien,Hossam M.Moftah,Ahmad TaherAzar,MahmoudShoman,” MRI diagnosis of breast cancer hybrid approach using adaptive ant-based segmentation and

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

multilayer perceptron neural networks classifier", *Applied Soft Computing*, Vol. 14,2014, pp. 62-71.

[7]. BichenZhengSang WonYoonSarah S.Lam, "Diagnosis of breast cancer based on feature extraction using a hybrid of K-means and support vector machine algorithms", *Expert Systems with Applications*, Vol. 41, 2014, pp. 1476-1482.

[8]. Mary C. Playdon Michael B. Bracken Tara B. Sanft Jennifer A. Ligibel Maura HarriganMelinda L. Irwin," Weight Gain After Diagnosis of breast cancer and All-Cause Mortality: Systematic Review and Meta-Analysis", *JOURNAL of the NATIONAL CANCER INSTITUTE*, Vol. 107, 2015.

[9]. AytuğOnan, "A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer ", *Expert Systems with Applications*, Vol. 42,2015, pp. 6844-6852.

[10]. Kriti Jitendra Virmani, Nilanjan Dey,Vinod Kumar, "PCA-PNN and PCA-SVM Based CAD Systems for Breast Density Classification" *Applications of Intelligent Optimization in Biology and Medicine*, Vol. 96, 2015, pp. 159-180.

[11]. Subrata KarEmail , D. Dutta Majumder, "An Investigative Study on Early Diagnosis of breast cancer Using a New Approach of Mathematical Shape Theory and Neuro-Fuzzy Classification System", *International Journal of Fuzzy Systems*, Vol. 18, 2016, pp. 349-366.

[12]. R Delshi Howsalya Devi1, Dr. M Indra Devi, OUTLIER DETECTION ALGORITHM COMBINED WITH DECISION TREE CLASSIFIER FOR EARLY DIAGNOSIS OF BREAST CANCER, *International Journal of Advanced Engineering Technology*, 2016. Pp. 1-6.

[13]. MehrbakhshNilashi,OthmanIbrahim,HosseinAhmadi,LeilaShahmoradi, A knowledge-based system for breast cancer classification using fuzzy logic method, *Telematics and Informatics*, Vol. 34, 2017, pp. 133-144.

[14]. Fatima S. Ahadi , Mark R. Desai , Chengwei Lei, Yu Li , Ruting Jia, Feature-Based classification and diagnosis of breast cancer using fuzzy inference system, *IEEE*, 2017.

[15]. LingrajDora,SanjayAgrawal,RutuparnaPanda,AjithAbraham, Optimal breast cancer classification using Gauss–Newton representation based algorithm, *Expert Systems with Applications*, Vol . 85, 2017, pp. 134-145.

[16]. Khan, S. U., Islam, N., Jan, Z., Din, I. U., Khan, A., & Faheem, Y. (2019). An e-Health care services framework for the detection and classification of breast cancer in breast cytology images as an IoMT application. *Future Generation Computer Systems*, *98*, 286-296.

[17]. Osmanović, A., Halilović, S., Ilah, L. A., Fojnica, A., & Gromilić, Z. (2019). Machine learning techniques for classification of breast cancer. In *World Congress on Medical Physics and Biomedical Engineering 2018* (pp. 197-200). Springer, Singapore.

Power System Technology
ISSN:1000-3673
Received: 06-06-2024 Revised: 15-07-2024 Accepted: 28-08-2024

[18]. Acharya, S., Alsadoon, A., Prasad, P. W. C., Abdullah, S., & Deva, A. (2020). Deep convolutional network for breast cancer classification: enhanced loss function (ELF). *The Journal of Supercomputing*, 1-18.

[19]. Wang, P., Song, Q., Li, Y., Lv, S., Wang, J., Li, L., & Zhang, H. (2020). Cross-task extreme learning machine for breast cancer image classification with deep convolutional features. *Biomedical Signal Processing and Control*, *57*, 101789.

[20]. Sujatha, K., Durgadevi, G., Kumar, K. S., Karthikeyan, V., Ponmagal, R. S., Hari, R., ... & Cao, S. Q. (2020). Screening and early identification of microcalcifications in breast using texture-based ANFIS classification. In *Wearable and Implantable Medical Devices* (pp. 115-140). Academic Press.

[21]. Sriraam, N., Murali, L., Girish, A., Sirur, M., Srinivas, S., Ravi, P., ... & Jeyanathan, J. (2020). Classification of Breast Thermograms Using Statistical Moments and Entropy Features with Probabilistic Neural Networks. In *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications* (pp. 1328-1340). IGI Global.

[22]. González-Patiño, D., Villuendas-Rey, Y., Argüelles-Cruz, A. J., Camacho-Nieto, O., & Yáñez-Márquez, C. (2020). AISAC: An Artificial Immune System for Associative Classification Applied to Breast Cancer Detection. *Applied Sciences*, *10*(2), 515.

[23]. Mushtaq, Z., Yaqub, A., Sani, S., & Khalid, A. (2020). Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets. *Journal of the Chinese Institute of Engineers*, *43*(1), 80-92.

[24]. Xu, Y., Zhu, Q., & Fan, Z. Q. (2013). Coarse to fine K nearest neighbor classifier. *Pattern recognition letters*, 980-986.

[25]. Malik, J. S., Goyal, P., & Sharma, A. K. (2010). A comprehensive approach towards data preprocessing techniques & association rules. In *Proceedings of the 4th National Conference*.

[26]. H. M. Zolbanin, D. Delen, and A. H. Zadeh, "Predicting overall survivability in comorbidity of cancers: A data mining approach," *Decision Support Systems*, vol. 74, pp. 150-161, 2015.