

ENSEMBLE LEARNING BASED FEATURE SELECTION WITH WEIGHTED AVERAGE LSTM NETWORK FOR BREAST CANCER RECURRENCE PREDICTION

¹Preetha G, ²Suban Ravichandran

¹Research Scholar, Department of Computer & Information Science, Annamalai University, India

²Associate Professor, Department of Information Technology, Annamalai University, India

Abstract

Breast cancer has recently become the subsequent most dangerous origin of cancer death in women, posing a serious threat to middle-aged women all over the world. However, early detection and prevention can greatly lower the risk of demise. Enhancing the likelihood of cancer repetition is a crucial aspect of the prognosis for breast cancer. For many researchers, predicting the return of breast cancer has been a difficult research topic. Increasingly, data mining techniques have drawn a lot of interest, particularly when they are applied to the creation of prognostic models using survival data. The prior system developed an effective weighted Minkowski radial basis function-based support vector machine and Simulated Annealing Inertia Weight based Chicken Swarm Optimization (SAIWCSO) method for breast cancer recurrence prediction (WMRBF-SVM). Large data sets, however, are not a good fit for the SVM method. Because when target classes coincide and the data set contains additional noise, SVM also performs poorly. The suggested approach created an ensemble learning-based feature selection using Weighted Average Long Short Term Memory Network (WALSTM) for prediction of breast cancer recurrence in order to address this issue. As an input, the Wisconsin Breast Cancer Dataset (WBCD) is used. The input data are preprocessed to eliminate irrelevant and/or missing value data. Using the z-score normalisation method, repetitive entries were removed during this phase. Following that, linear function-based animal migration optimization and ensemble learning-based feature selection are carried out (LFAMO). Classification is carried out utilizing the Weighted Average Long Short-Term Memory Network according to the selected attributes (WALSTM). Utilizing MATLAB, the experiments are reproduced. According to the experimental findings, the suggested system performs better than the current system in terms of accuracy, precision, recall, and specificity, f-measure.

Keywords: Breast cancer, ensemble-based feature selection, Information Gain (IG), Animal Migration Optimization (AMO) and Weighted Average Long Short Term Memory Networks (WALSTMs).

1. INTRODUCTION

Breast cancer is one of the most common cancers in women globally [1]. The second most frequent cancer to be diagnosed worldwide, it accounted for 6.4% of all fatalities, and that it was the fifth greatest cause of cancer mortality globally, according to GLOBOCAN's most recent statistics. With 521 907 recorded fatalities in 2012, it is linked to the greatest rate of cancer-related deaths among women. Breast cancer can affect men as well, despite it

typically affecting women. Male breast cancer, however, is uncommon; it accounts for fewer than 1% of all instances. The occurrence of aberrant cells growing out of control is known as cancer. However, many patients with breast cancer may not exhibit severe symptoms and/or may attribute fatigue and weight loss, which are potential cancer signs, to a variety of unrelated factors (stress, a change in diet, insufficient sleep) [2-3]. Breast lumps or general alterations are two examples of external indications of BC. The possibility of cancer emerges when a patient finds a breast anomaly (via self-examination or during a doctor visit) or after a mammogram indicates it.

Recurrence happens when cancer reappears a short time after treatment [4]. Early cancer detection and prognosis have become essential for clinical management and to help with hospital attention in cases of breast cancer recurrence. Data mining helps to manage the growing amount of medical data and provides helpful medical advancement and malignant disease treatment as medical technology advances. The application of various machine learning approaches can help doctors make decisions that are precise and effective.

Two crucial elements in the prediction of breast cancer are feature selection and categorization. The process of choosing features that are pertinent to the classification process is known as feature selection. The goal of this is to enhance classification performance outcomes. The learning process can be accelerated via feature selection by removing some useless data. For the best feature selection, numerous optimization methods have been used in the past, including Particle Swarm Optimization (PSO) [5–6], Genetic Algorithm (GA), Firefly Algorithm (FA), and Ant Colony Optimization (ACO). According to some study on the subject, artificial intelligence (AI) or machine learning (ML) can be used to diagnose breast cancer sooner thanks to technological improvements [7]. It is unable to however, deliver results for classification accuracy that are acceptable. To increase the accuracy and true positive rate, better feature selection and machine learning techniques are therefore needed.

The remnants of the essay is systematized as follows. A survey of the many breast cancer recurrence prediction techniques is presented in Section 2. The suggested Ensemble learning based feature selection using the Weighted Average Long Short-Term Memory Networks (WALSTMs) model is described in Section 3. Compare the effectiveness of the suggested strategy with the current methods in Section 4. The last section of this essay is Section 5.

Recently, Data mining has become a popular and efficient tool for knowledge discovering and extracting hidden patterns from large datasets. It involves the use of sophisticated data manipulation tools to discover previously unknown, valid patterns and relationships in large dataset. We applied three strong data mining classification algorithms i.e. SVM, Naive Bayes and C4.5 Decision Tree on a medium sized dataset which contained 35 attributes and 198 cancer patient data. We classified the instances on the basis of 'output' attribute which can have only two nominal values i.e. 'R' for recursive cancer and 'N' for non-recursive cancer. Results have shown that SVM has higher prediction accuracy i.e. 75.75 % than Naive Bayes (67.17 %) and C4.5. Recently, Data mining has become a popular and efficient tool for

knowledge discovering and extracting hidden patterns from largedatasets. It involves the use of sophisticated data manipulation toolsto discover previously unknown, valid patterns and relationships in large dataset. We appliedthree strong data mining classification algorithms i.e. SVM, Naive Bayes and C4.5 Decision Tree on a medium sized dataset which contained 35 attributes and 198 cancer patient data. We classified the instances on the basis of 'output' attribute which can have only two nominal values i.e. 'R' for recursive cancer and 'N' for non-recursive cancer. Results haveshown that SVM has higher prediction accuracy i.e. 75.75 % than Naïve Bayes (67.17 %) and C4.

2. LITERATURE REVIEW

With the SVM and KNN for breast cancer exposure by training its possessions, [9] the researcherprovides a novel method for the prediction of same. The solution that is being presented uses 10-fold cross validation to produce truthful results. The UCI machine learning repository was used to obtain the WBCD. The strategy yields superior training and testing results. Additionally, the methods in the testing phase had specificities of 95.65% and 92.31% and accuracy of 98.57% and 97.14% for Support Vector Machine and K-Nearest Neighbors, respectively [9].

In a 2010 study, Strumbelj et al. revealed the first practical application of two strategies to communicate more effectively. The initial is a technique for explaining classifier forecasts that provides additional information to the user regarding how a classifier chooses what and how to suggest. The second is a mechanism for estimating the reliability of regression forecasts, which aids users in determining how much confidence to place in a given prediction. Both approaches are effectively used to a brand-new dataset for the estimation of diseased recurrence, and the outcomes are assessed by leading doctors [10].

New techniques were revealed by Pritom in 2016. The WBCD was used to get the data for the cancer patient. The prediction accuracy of the NB, C4.5 DT, and SVM classification algorithms was calculated for the dataset's total of 35 attributes. We were able to increase the accuracy by an effective feature selection technique. These features not only make very little contributions, but their incorporation also confuses the categorization algorithms. All three algorithms' accuracy rates were significantly increased after a thorough selection of higher rated features [11]. Use resampling strategies next to address the class imbalance problem. Finally, to forecast the likelihood of recurrence, develop two classifiers: AdaBoost and cost-sensitive learning. Achieve accuracy of 0.973 and sensitivity of 0.675 by using the AdaBoost algorithmwhich virtually eliminates incorrect dismissals. [12].

In order to forecast the likelihood of ainfection recurrence inside the Korean population following 5 years of surgery they developed a unique prognostic model based on SVM [13]. Retrospective data on 679 individuals who endured surgery for breast cancer were gathered. The construction and comparison of three prediction methods using the SVM, ANN, and Cox-proportional hazard regression model. The SVM (BCRSVM) model used in this study beat previous prognostic models (area under the curve for the BCRSVM was 0.85, 0.71, and 0.70, respectively). The BCRSVM demonstrated significantly high efficiency among other metrics [13].

3. PROPOSED METHODOLOGY

For the diagnosis of recurrence, the suggested system developed an ensemble learning-based feature selection method using Weighted Average Long Short Term Memory Networks (WALSTMs). Data preparation, data normalisation, feature selection, and classification are the four primary stages that make up this procedure. Figure 1 displays the planned work's flowchart.

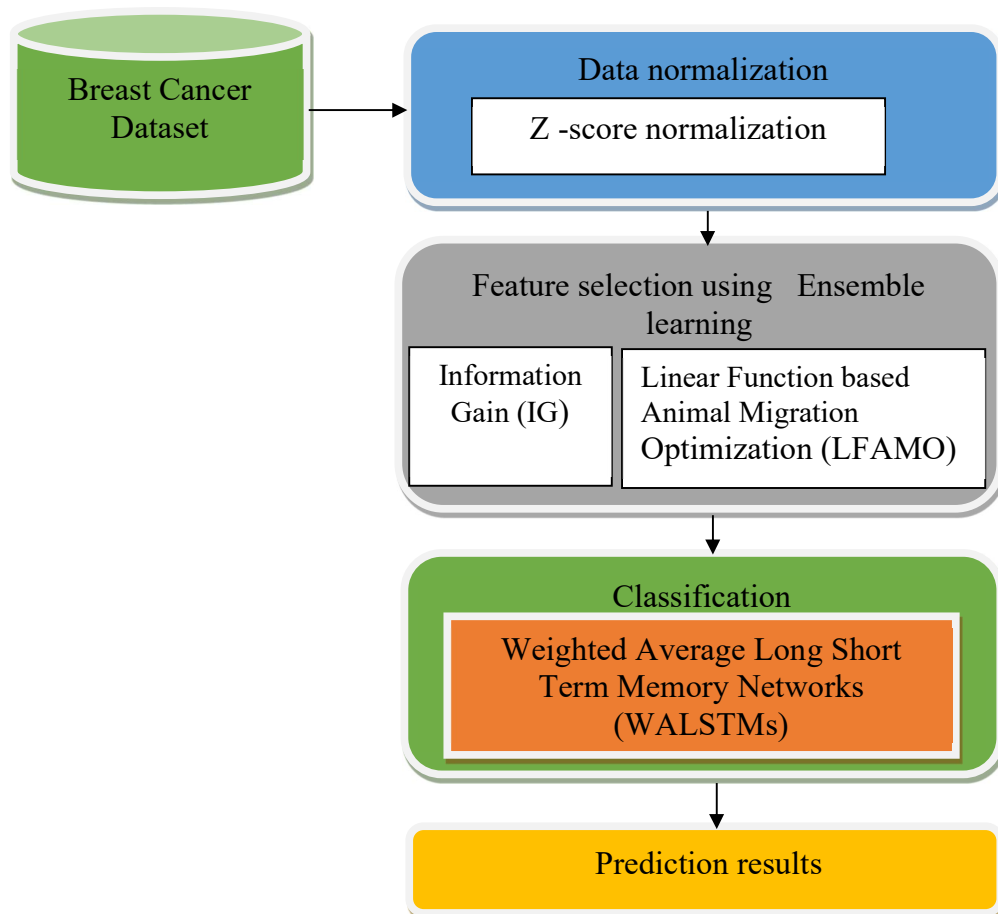


Figure 1: Flow diagram of the proposed work

3.1 Input and Data Normalization

As an input, the WBCD is used. The input data are preprocessed to eliminate irrelevant and/or missing value data. Data preparation is a crucial step in the progression of knowledge discovery. It is regarded as the cornerstone of data mining. A subset of data analysis known as "data normalisation" involves scaling attribute data or feature values to lie within a predetermined range, such as -1.0 to 1.0 or 0.0 to 1.0. The most used normalisation technique is Z-score normalisation. Zero mean normalisation is another name for Z score normalisation. The data in this instance has been standardised using the mean and standard deviation. then the equation is,

$$d' = \frac{d - \text{mean}(p)}{\text{std}(p)} \quad (1)$$

Where,

Mean(p) -is the average of all p's attribute values.

Std(p)-is the standard deviation of all p's value.

3.2 Feature selection

The persistence of feature selection techniques is to discard irrelevant or superfluous features from a given feature vector. The Ensemblelearning based feature selection is done based on Information Gain (IG) and Linear Function based Animal Migration Optimization (LFAMO) algorithm. Ensemble techniques are initially presented to cope with the steadiness of feature selection technique.

A. Information Gain (IG)

To determine which characteristic offers the most information about a classification based on the concept of entropy, Information Gain (IG) is used. Information Gain (IG) calculations heavily rely on the idea of entropy [14]. Using the distinctions amongst entropy and conditional entropy, or IG, it is possible to assess the significance of features inside a given category. The reduction of uncertainty as defined below is indicated by IG $g(Y, X)$.

$$g(Y, X) = H(Y) - H(Y|X) \quad (2)$$

While $H(Y|X)$ signifies the conditional entropy, which signifies the indecision derived from the known variable, $H(Y)$ signifies the randomness of dataset Y, which enumerates the indecision associated in forecasting the price of a random variable. P stands for likelihood distribution X. Then

$$H(Y) = -\sum p(y) \log p(y) \quad (3)$$

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x) \quad (4)$$

Each feature is scored conferring to its IG value, and the highest-ranked features are chosen as the best features.

B. Linear Function based Animal Migration Optimization (LFAMO) algorithm

The AMO algorithm was created to handle global optimization issues and was motivated by the animal migration patterns that can be observed in all foremost animal groups, including fish and birds. The fundamental operation of the algorithm is described in terms of two idealized assumptions: (1) The herd's leader animal will be determined by the animal with the highest quality, and the leader animal will indeed be preserved for future generations. (2) The herd has a set number of animals, and each one will be substituted with a novel animal with likelihood P_a . In this situation, they will quit the group and later be replaced by a new animal [15–16].

The migration process and the population update method make up the AMO algorithm. The animal's transition from one location to another is covered by the migration process. The following list of animal herds makes up the migrating animal population. Eq displays this migrating population (5),

$$\text{Population} = \{X_1, X_2, \dots, X_{NP}\} \quad (5)$$

Where, X_{NP} and X_i , respectively, stand for the dataset's population and feature sizes. Each feature in the dataset is made up of a d-dimensional vector that has been acknowledged as a component of a solution falling inside the search space's maximum and lowest bounds. The algorithm starts with a population that is initialized randomly and is expressed as follows:

$$X_i = X_{min} + rand * (X_{max} - X_{min}) \quad (6)$$

Where, X_i , X_{min} and X_{max} stand for a population feature, the minimum and maximum boundaries of the search space, and a feature in the population, respectively. A "rand" is a random number between 0 and 1 that has a uniform distribution.

1. Practice of Animal Relocation

An animal should adhere to three rules when migrating: (1) stay clear of your neighbours; (2) travel in the same direction as your neighbours; and (3) stay near to your neighbours. We employ a topological ring to define the idea of an individual's immediate neighbourhood. We chose a neighbourhood length of five for each of the various dimensions to keep things simple. Be aware that our technique defines the neighbourhood topology on a collection of vector indices and that this definition is static.

When a feature's index is 1, its neighbourhood is made up of features with indices $NP - 1$, NP , 1, 2, 3, and so on. If the feature's index is I then its neighbourhood is made up of features with indices $i - 2$, $i - 1$, i , $i + 1$, $i + 2$. We select one neighbour at random from the neighbourhood topology after which we modify the person's location to reflect this neighbour, as indicated by the given equations:

$$X_i^{t+1} = X_i^t + \delta \cdot (X_{neighbor}^t - X_i^t) \quad (7)$$

$$\delta = \delta_{max} - \frac{\delta_{max} - \delta_{min}}{iter_{max}} \times k \quad (8)$$

Where,

δ_{min} and δ_{max} - Lower and upper bounds of the permissible range of δ

$iter_{max}$ -Maximum number of iterations

k -Constant

Where $X_{neighbor}^t$ is the neighborhood's current location, X_i^t and X_i^{t+1} are the i th person's positions in iterations t and $t + 1$, respectively, and is a random number that can be improved using a straightforward linear function. X_i^t is the present spot of the i th individual, and X_i^{t+1} is the new location of the i th individual, which increases the diversity of animal populations in the following search phase.

2. Population Updating Process

Even throughout this update process, the computer duplicates how some creatures quit the group and a few migrate to the new community. Each member of the population receives a probability value based on their fitness value. Classification accuracy is taken into account as a fitness function in the suggested study. While the probability value for the population's most

compatible member is 1, the probability value for its most mismatched member is $1/NP$. When the probability value of the separate is lower than the value that was determined at random, Eq.

$$X_i^{t+1} = X_{r1}^t + rand * (X_{best}^t - X_i^t) + rand * (X_{r2}^t - X_i^t) \quad (9)$$

Where, X_i^t and X_i^{t+1} are, respectively, the i th feature's locations in iterations t and $t+1$. X_{r1}^t and X_{r2}^t are randomly chosen features from the population, X_{best}^t are the features that are in the top positions and have excellent quality, and $rand$ is a random value between 0 and 1. If the new individual has higher quality than the existing individual, the existing individual is eliminated from the population and the new individual is added.

Algorithm 1: LFAMO

1. Initiate
2. Set the reproduction counter $G=0$ and set the NP individuals X_i at random.
3. analyse each person's level of fitness
4. the terminating criteria are not met, proceed
5. for $i = 1$ to NP do
6. for $j = 1$ to D do
7. $X_i^{t+1} = X_i^t + \delta \cdot (X_{neighbor}^t - X_i^t)$
8. end for
9. end for
10. for $i = 1$ to NP do
11. examine the patient's fitness X_i^{t+1}
12. If X_i^{t+1} is healthier than X_i then
13. $X_i = X_i^{t+1}$
14. End if
15. end for
16. for $i = 1$ to NP
17. for $j = 1$ to D
18. select $r1$ and $r2$ randomly ($r1 \neq r2 \neq i$)
19. if $rand > Pa$ then
20. $X_i^{t+1} = X_{r1}^t + rand * (X_{best}^t - X_i^t) + rand * (X_{r2}^t - X_i^t)$
21. end if
22. end for
23. end for
24. for $i = 1$ to NP do
25. evaluate the fitness of the offspring X_i^{t+1}
26. If X_i^{t+1} is better than X_i then
27. $X_i = X_i^{t+1}$
28. end if
29. end for
30. keep in mind the most successful outcome thus far
31. end while

32. end

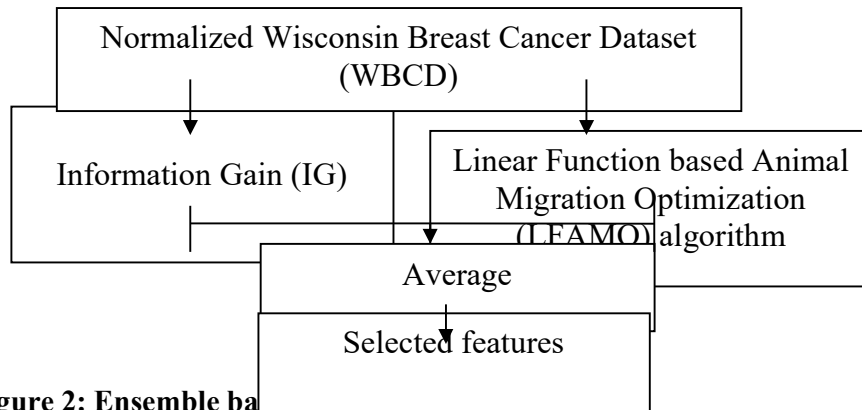


Figure 2: Ensemble ba

Before performing the feature selection task, for the provided input, the output probabilities are averaged, which are from each algorithm. i.e. Information Gain (IG) and Linear Function based Animal Migration Optimization (LFAMO) algorithm. For a given input features, the output i of algorithm j is represented by $r_j(i)$.

$$S_i = \frac{1}{n} \sum_{j=1}^n r_j(i) \quad (10)$$

3.3 Classification using WALSTMs

Based on the selected features, classification is performed by using WALSTMs. The LSTM is a RNN architecture which is formed to highly correctly simulate temporal sequences and their long-range dependencies than standard RNNs. As illustrated in Figure 3, a standard LSTM cell consists of input, forget, output gates, as well as cell activation element [17]. Such units obtain activation signals from many sources and use specified multipliers to control cell activation. For many time steps, the LSTM gates could avoid the remainder of network from changing content of memory cells. Signals are preserved and errors are propagated for considerably longer in LSTM network than in RNNs.

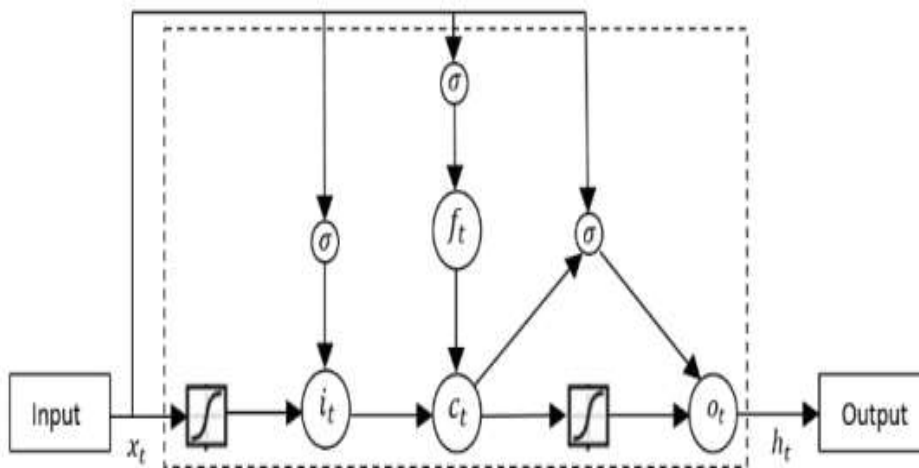


Figure 3. LSTM Cell

LSTM input gate is described as

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (11)$$

The forget gate is defined as

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (12)$$

The cell gate is defined as

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (13)$$

The output gate is defined as

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (14)$$

Lastly, hidden state is calculated as

$$h_t = o_t \tanh(c_t) \quad (15)$$

tanh - hyperbolic tangent activation function

x_t - input at t time

W and b - network parameters (Weights, Biases)

The weight value is updated based on the weighted average scheme. When this is the case the system can apply a weighting to each value. The weighted mean can be calculated with the subsequent formula:

$$W = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (16)$$

Where, x_i is the features, w_i is a corresponding weight for each feature, σ is represented by logistic sigmoid function, i, f, o and c are input, forget, output, cell state, correspondingly. Weight matrices for peephole acquaintances are indicated by letters W_{ci} , W_{cf} and W_{co} . Three gates in an LSTM evaluate the communications (i, f, o). The input gate chooses the input ratio. This ratio has an impact on the equation when defining the cell state (13). The forget gate verifies if prior memory h_{t-1} is passed or not. The prior memory's ratio is determined in equation (12) and utilized in the equation (13). The output gate decides if memory cell's outcome is passed or not. This is seen in equation (14). Addition to the three gates, they can address the

vanishing and exploding gradient problems with LSTM. Recurrent hidden layer is substituted with an LSTM cell in the LSTM-RNN architecture. The output gate classifies the dataset into two categories such as cancer or normal.

4. EXPERIMENTAL RESULTS

The performance of the proposed work is evaluated using python. The dataset is collected from <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>. The proposed Weighted Average Long Short Term Memory Network (WALSTM) is compared with previous Weighted Minkowski Radial Basis Function based Support Vector Machine (WMRBF-SVM), Naive Bayes (NB) and fast decision tree learner (REPTree) and approaches in terms of accuracy, precision, recall, f-measure and specificity. Table 1 represents the overall performance comparison.

4.1 Accuracy

Accuracy is the most correct performance estimation that accurately distinguishes an event and is simply a prediction of the right predictions in all observations.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (17)$$

Where,

TP - True Positive

FN - False Negative

FP - False Positive

TN - True Negative

4.2 Precision

Precision defines the relevance of the results and is given by the ratio positive observations predicted properly to the positive observations predicted in total.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (18)$$

4.3 Recall

Recall is the ratio of properly estimated positive observations to the all observations in actual class - yes.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (19)$$

4.4 F-measure

F1 Score is also a measure of accuracy of experiment and is defined by the weighted mean of Precision and Recall.

$$\text{F-measure} = 2 * \frac{(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (20)$$

4.5 Specificity

Specificity of a classifier is the ratio between how much were properly classified as negative to how much was actually negative.

$$\text{Specificity} = \frac{TN}{FP+} \quad (21)$$

Table 1: Performance table

Metrics	Methods			
	WALSTM	WMRBF-SVM	NB	REPTree
Accuracy	0.97	0.85	0.81	0.80
F1-score	0.92	0.75	0.61	0.46
Precision	0.90	0.87	0.60	0.63
Recall	0.93	0.619	0.63	0.37
Specificity	0.91	0.90	0.63	0.36

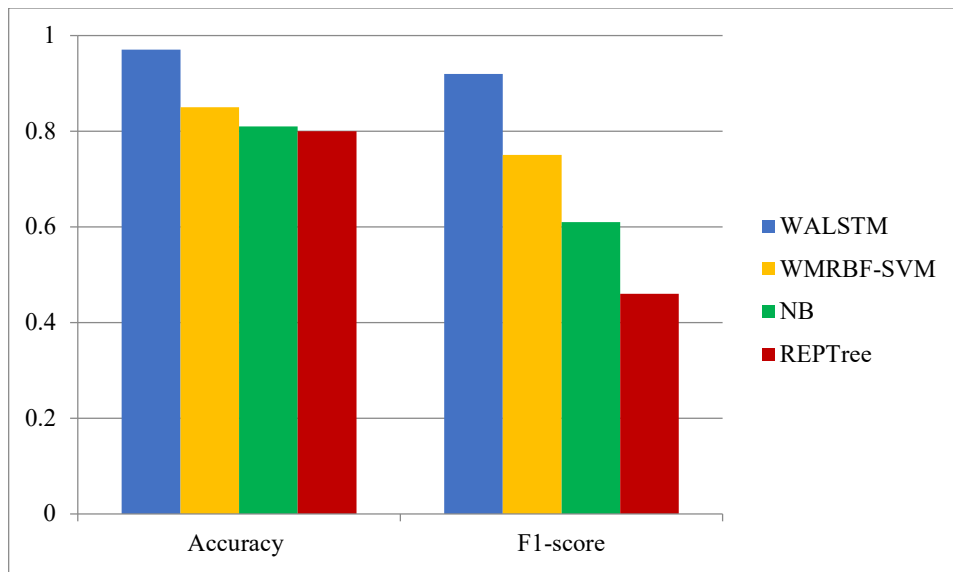


Figure 4 : Accuracy and F1-score comparison

Accuracy and F1-score of the Weighted Average Long Short Term Memory Network (WALSTM) is compared with previous Weighted Minkowski Radial Basis Function based Support Vector Machine (WMRBF-SVM), Naive Bayes (NB) and fast decision tree learner (REPTree). The proposed WALSTM attains accuracy of 0.97 and F1-measure of 0.92 which are better than previous WMRBF-SVM, NB and REPTree methods.

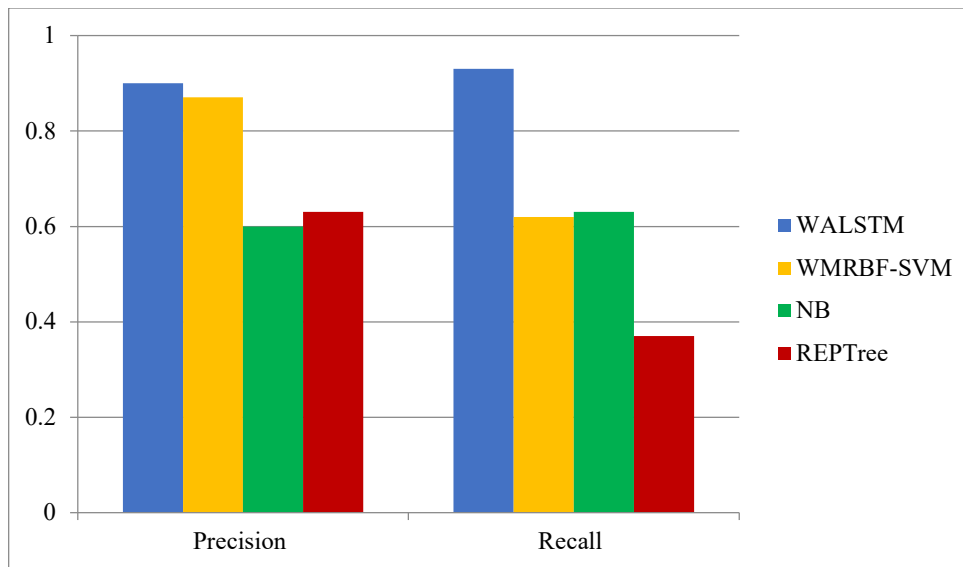


Figure 5: Precision and Recall comparison

The Weighted Average Long Short Term Memory Network (WALSTM) is compared with previous Weighted Minkowski Radial Basis Function based Support Vector Machine (WMRBF-SVM), Naive Bayes (NB) and fast decision tree learner (REPTree) interms of precision and recall. In this work, the ensemble learning based feature selection is done based on Information Gain (IG) and Linear Function based Animal Migration Optimization (LFAMO) for optimal feature selection. The proposed WALSTM attains precision of 0.90 and recall of 0.93 which are better than previous WMRBF-SVM, NB and REPTree methods.

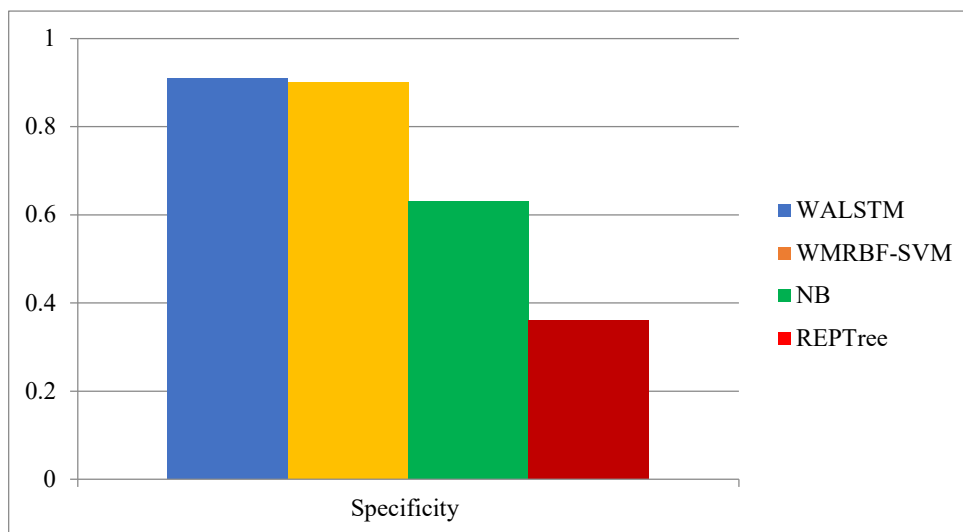


Figure 6: Specificity comparison

Specificity of the Weighted Average Long Short Term Memory Network (WALSTM) is compared with previous Weighted Minkowski Radial Basis Function based Support Vector

Machine (WMRBF-SVM), Naive Bayes (NB) and fast decision tree learner (REPTree). To improve the classification accuracy, the classification is performed by using WALSTMs. The proposed WALSTM method attains 0.91 of specificity when other methods such as WMRBF-SVM, NB and REPTree attain 0.90, 0.63 and 0.36 respectively.

5. CONCLUSION

The proposed system designed an Ensemble learning based feature selection with WALSTM for accurate breast cancer recurrence prediction. To normalize the data in the dataset, Z score normalization technique is used. In order to improve the classification accuracy, Ensemble based feature selection which includes Information Gain (IG) and Linear Function based Animal Migration Optimization (LFAMO) algorithm is utilized for optimal feature selection. At last, Weighted Average Long Short Term Memory Network (WALSTM) scheme is introduced for breast cancer repetition forecast. According to the experimental findings, the new system performs better than the old system in terms of accuracy, precision, recall, specificity and f-measure.

REFERENCES

- [1] Abreu, P. H., Santos, M. S., Abreu, M. H., Andrade, B., & Silva, D. C. (2016). Predicting breast cancer recurrence using machine learning techniques: a systematic review. *ACM Computing Surveys (CSUR)*, 49(3), 1-40.
- [2] Ziner, K. W., Sledge Jr, G. W., Bell, C. J., Johns, S., Miller, K. D., & Champion, V. L. (2012, May). Predicting fear of breast cancer recurrence and self-efficacy in survivors by age at diagnosis. In *Oncology nursing forum* (Vol. 39, No. 3, p. 287). NIH Public Access.
- [3] Brothers, B. M., & Andersen, B. L. (2009). Hopelessness as a predictor of depressive symptoms for breast cancer patients coping with recurrence. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer*, 18(3), 267-275.
- [4] Ahmad, L. G., Eshlaghy, A. T., Poorebrahimi, A., Ebrahimi, M., & Razavi, A. R. (2013). Using three machine learning techniques for predicting breast cancer recurrence. *J Health Med Inform*, 4(124), 3.
- [5] Agustian, F., & Lubis, M. D. I. (2020, October). Particle swarm optimization feature selection for breast cancer prediction. In 2020 8th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-6). IEEE.
- [6] Sakri, S. B., Rashid, N. B. A., & Zain, Z. M. (2018). Particle swarm optimization feature selection for breast cancer recurrence prediction. *IEEE Access*, 6, 29637-29647.
- [7] Shukla, R., Yadav, V., Pal, P. R., & Pathak, P. (2019). Machine Learning Techniques for Detecting and Predicting Breast Cancer. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(7), 2658-2662.
- [8] 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), pp. 687-699.

- [9] Md. Milon Islam, Hasib Iqbal, Md. Rezwanul Haque, and Md. Kamrul Hasan, "Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors", 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017.
- [10] Štrumbelj, E., Bosnić, Z., Kononenko, I., Zakotnik, B., & GrašičKuhar, C. (2010). Explanation and reliability of prediction models: the case of breast cancer recurrence. *Knowledge and information systems*, 24(2), 305-324.
- [11] Pritom, A. I., Munshi, M. A. R., Sabab, S. A., & Shihab, S. (2016, December). Predicting breast cancer recurrence using effective classification and feature selection technique. In *2016 19th International Conference on Computer and Information Technology (ICCIT)* (pp. 310-314). IEEE.
- [12] Yang, P. T., Wu, W. S., Wu, C. C., Shih, Y. N., Hsieh, C. H., & Hsu, J. L. (2021). Breast cancer recurrence prediction with ensemble methods and cost-sensitive learning. *Open Medicine*, 16(1), 754-768.
- [13] Kim, W., Kim, K. S., Lee, J. E., Noh, D. Y., Kim, S. W., Jung, Y. S., ... & Park, R. W. (2012). Development of novel breast cancer recurrence prediction model using support vector machine. *Journal of breast cancer*, 15(2), 230-238.
- [14] Lei, S. (2012, March). A feature selection method based on information gain and genetic algorithm. In *2012 international conference on computer science and electronics engineering* (Vol. 2, pp. 355-358). IEEE.
- [15] Li, X., Zhang, J., & Yin, M. (2014). Animal migration optimization: an optimization algorithm inspired by animal migration behavior. *Neural Computing and Applications*, 24(7), 1867-1877.
- [16] Lai, Z., Feng, X., & Yu, H. (2019, May). An Improved Animal Migration Optimization Algorithm Based on Interactive Learning Behavior for High Dimensional Optimization Problem. In *2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)* (pp. 110-115). IEEE.
- [17] Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. *Artificial Intelligence Review*, 53(8), pp.5929-5955.