International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Identifying The Best-Fit Associative Classifier For Determining Survivability And Non-Survivability In Breast Cancer Patients

Ms. Sheethal Aji Mani [1], Dr. Thivakaran T. K [2]

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract: In developing nations, cancer death is considered one of the biggest challenges. Even though there are various strategies to prevent cancer, some types of cancer still receive inadequate treatment. One of these is breast cancer. Early diagnosis is very important in treating this disease. Breast cancer has a good survival rate, especially when it is detected early. This can be attributed to better treatment and early diagnosis. This study aims to classify the various medical attributes collected from the SEER dataset for breast cancer patients into two categories: non-survivability and survivability. The paper used associative classifiers such as ACAC, ACN, L3, and CBA2 to analyze the data. The objective of the study is to identify a best fit classifier for deciding survivability and non-survivability in breast cancer patients based on accuracy. The outcome of the study revealed that CBA and CBA2 model exhibited an accuracy of 81% even in a huge dataset with 44325 records. The proposed approach also showed an improvement in performance. These findings indicated that the possibility of identifying non-survivability and survivability in breast cancer patients could be explored

Keywords: *Classification, Breast cancer, Survivability, Associative Classifier*

1. Introduction

Data mining involves identifying patterns in large sets of information. The adoption of KDD methods has increased significantly in the recent past years due to the evolution of warehousing technology. This helps one to derive useful knowledge from raw information.

Decision-making process can be improved by identifying and predicting the outcomes of their data. The techniques used in this process can be categorized into two main categories: describe the target data and predict its outcomes using machine learning. They are used to filter and organize the data, identify the most interesting pieces of information, identify security breaches and fraud, detecting diseases like cancer etc.

A wide variety of datasets [1] can be utilized in different industries like marketing, healthcare, and social media. Unfortunately, only a few of them are interpreted by data scientists and are considered sufficient for making predictions. Due to the increasing number of marketers who are starting to analyze their data, they are able to make more informed decisions and enhance their effectiveness. This exercise can help them implement new strategies and perk up their efficiency [2].

One of the prevalent types of cancer among women worldwide is breast cancer [3]. According to global statistics, it is responsible for the majority of new cases and deaths due to cancer related issues. This condition is a crucial community health concern.

Early diagnosis of breast cancer can lead to a better prognosis and increase the chances of survival. Further analysis can help prevent patients from going through unnecessary treatments. This is the reason why it is important that the correct classification of breast cancer is carried out. Machine learning [4] is conventional in the field of breast carcinoma prognosis and pattern classification due to its ability to detect important features of complex datasets.

In data mining, the classification [5] process involves assigning a label to every instance of a dataset based on its attributes. The intent of this process is to evolve a classification model that can precisely predict the changes in class labels in new instances. Data mining and classification techniques are widely used in the field of medical analysis and diagnosis to make informed decisions.

The SEER dataset was used in this study to classify breast cancer patients. The algorithms used in this process included ACAC, ACCF, ACN, L3, CBA, and CBA2. The public data included 14 attributes that are known to have a notable impact on a patient's survival chances. The research goal was to develop a model that can accurately predict the survival chances of patients with breast cancer. The model's strength and the predictive factors were then compared with accuracy and precision measures.

[1] *Research Scholar, Dept. of Computer Science & Engineering, Presidency University, Bangalore*
sheethalabraham@gmail.com
ORCID:0009-0008-7342-6615

[2] *Professor & Dean-IQAC, Dept. of Computer Science & Engineering, East Point Group of Institutions, Bangalore*
tktpresidency4@gmail.com

The organization of the rest of the study is arranged as follows. The section II of the paper reviews the various ideas of the associative approach. Section III focuses on the proposed methods and their application to a given dataset. Section IV presents the performance study's results and the experimental outcomes. The paper concludes in section V with the final chapter.

2. Literature Review

Due to the variety of information collected and stored in hospitals, the healthcare and wellness domain is regarded as one of the most appropriate environments for data science. This field is also an ideal candidate for implementing data mining and machine learning techniques. Several studies have been done on the classification of breast cancer data. These findings show that most of them have an acceptable accuracy [6].

Chaurasia et al. [7] utilized a Wisconsin dataset and various supervised training methodologies such as neural networks, SVM and decision tree methods to analyze the performance of the classifiers. They found that SVM performed well and had 96.84 percent of accuracy value.

Delen et al. [8] analyzed the data gathered for the study, which involved over 200,000 breast cancer records, and categorized it into two groups: one composed of survivors, and another was not. The algorithms that were used for the classification were: the neural network, nave Bayes, and the c4.5 decision tree. The results of the evaluation revealed that the c4.5 decision trees was more acceptable than the other methods.

Srinivas et al. [9] utilized the two algorithms for the study, which involved making predictions about heart attacks based on various medical profiles, such as blood pressure, gender , age, and sugar levels. The results of the analysis revealed that the nave Bayes model performed better than the nave creedal model.

Bernal et al. [10] used a variety of methods to analyze the data collected for the study, such as neural networks, logistic regression, and k-nearest neighbors. They were able to improve the accuracy of their results by detecting the changes in the patients' conditions over a 24-hour period. They were able to achieve an accuracy with k-nearest neighbor method and logistic regression. According to Bernal, the researchers should consider the various parameters when it comes to developing a model.

In a study performed by Williams et al. [11] , they looked into the use of data mining techniques to predict the likelihood of breast cancer in women in Nigeria. This disease is prevalent among women in this country, and there are limited services that can help predict it. To find an efficient method to help women with this disease, the researchers utilized the decision tree J48 and naiive Bayes.

Oyewola et al. [12], conducted a study on the prediction of breast cancer based on a mammographic diagnosis. They utilized various methods such as the use of logistic regression, linear discriminant analysis, and random forest and SVM. The results revealed that Support Vector Machine model provides highest accuracy for the prediction with mammography.

Aruna et al. [13] utilized decision trees, nave Bayes, and support vector machines to classify the data collected from Wisconsin regarding breast cancer. They were able to achieve an accuracy of 96.99% with the Support Vector Machine model.

Quan et al. [14] used the results of the evaluation to determine which algorithm was the best when it came to classifying the diabetic patients. After comparing the various methods, the researchers concluded that the nave Bayes was the most accurate when it came to identifying diabetic patients with an accuracy rate of 80.8%.

In a study conducted by Asri and colleagues [15], they evaluated the performance of various machine learning algorithms. The four algorithms included the SVM, naiive Bayes, k-nearest neighbour, and decision tree. The goal of the evaluation was to analyze the efficiency and effectiveness of each algorithm by comparing its sensitivity, precision, and specificity. After the evaluation, the researchers revealed that Support Vector Machene had the highest score at 97.13%.

Wang et al. [16] used the collected data to find the most effective method for predicting breast cancer. They utilized various methods such as the use of neural networks, decision tree, and support vector machines. They then applied a principle component analysis method to analyze the data and reduce its features. The researchers evaluated the models' performance by using two datasets, one of which was the Wisconsin Breast Cancer database[17,18]. They also provided a comprehensive evaluation of the errors they encountered.

According to Nithya et al. [19], the main issue regarding breast cancer is the classification of the tumor. Currently, CAD is utilized for the diagnosis and study of this disease. Their objective was to utilize data mining approaches to enhance the prediction of breast cancer. They utilized various methods such as SVM-SMO, multilyer perceptron, multiboot, bagging, random subspace allocation to the naiive bayes performance.

Agarap et al. [20] used various methods such as the use of SVM, KNN, multilayer perceptron, GRU-SVM and softmax regression. The most accurate performance was shown by the multilayer perceptron, which had a 99.4% accuracy score.

The concept of an association-based classification system seeks to create compact, accurate, and efficient models by combining the expertise of association mining and classification. According to studies, this approach can provide a more accurate and compact classification system than the traditional methods. Jiri et al. [21] analyses the two versions of CBA(M1 and M2)Classification based on Association models and revealed that CBA version 1 is more faster in most of the scenario.

Huang et al. [22] proposes another associative classifier ACAC (All-Confidence based Associative Classification) where the support and all confidence measure is used to mine the data for multiple itemsets. The ACAC method then produces a small set of quality rules, which can be used to categorize new objects. The researchers then use the numerical and average information about the rules to measure the effectiveness of the group rules.

In another study [23] Li et al. introduce the ACCF(Associative Classification Based on Closed Frequent Itemsets) method, which is an associative method that takes into account the frequent itemsets. It enables them to create Class Association Rules by mining all the CFIs(Closed Frequent Itemsets) and their tidset. The researchers found that ACCF is more accurate and consistent when it comes to creating classification models in various databases compared to CBA.

Kundu et al. [24] suggested another model to provide a comprehensive analysis of the various issues that arise when it comes to creating accurate and efficient classification models. It aims to solve these issues by developing a robust and efficient method for creating negative associations. This process can be performed efficiently by using the ACN (Associative Classifier with negative rules) method and ensuring that the number of high-quality rules is sufficient.

A new associative classification method is introduced in this study conducted by Baralis et al. [25]. It uses a lazy pruning technique to discard the rules that only produce wrong case classifications. The first step of the process is to evaluate the existing rules that have correctly classified a training case. The remaining rules that are not used during the training phase are then sorted by confidence. The results revealed that the classification precision has been improved compared to previous models.

3. Methods and Application

3.1 Associative Classification

The concept of associative classification combines the expertise of data mining and association rule generation [32,33]. It allows researchers to create a set of association rules that are focused on the right-hand side of the association. After a large number of rules are generated, the researchers use a combination of rule ranking and pruning to select a subset of high-quality ones [34,35]. The reduced set of association rules will be used to create effective classification models. Compared to the decision tree method or any traditional classifiers, associative classification is more accurate [26,27,28].

The main principle of AC is that before it can be applied to a dataset, it has to be preprocessed. This process involves taking out the superfluous attributes and instances from the data and handling missing values. After that, association based rule extraction is performed to mine the frequent itemsets. The collected itemsets are used as input to create a set of candidate rules. These rules are then built into a decision tree, which is recursively split into smaller groups according to the attributes of the data.

After building a decision tree, using cost complexity or reduced error pruning techniques, the unneeded rules will be removed. They then use the built-in step three to classify new instances. The algorithm takes into account the leaf nodes' attributes to arrive at a classification label.

3.2. CBA Classifier (Classification Based on Association)

The classification algorithm is composed of two phases: the association rule generation phase (CBA – RG) and the classifier building phase(CBA-CB). In the association rule generation phase, the researchers find all the necessary rules that satisfy the minimum support and confidence levels. The latter is responsible for eliminating the repeating rules and building a new classification label. The speed at which the CBA algorithm was able to perform in a benchmark was one of the main advantages it had[29].

The first step in the classification algorithm's development is to generate a list of potential classification labels (CRs) according to the given schema.

If confidence value of rule r1 is greater than confidence of rule r2, then rule r1 is prioritized over rule r2.

If confidence in rule r1 is to that of rule r2 , then the priority of rule r1 is higher than that of rule r2 if support value of rule 1 is greater than that of rule 2.

The size of the antecedent is considered when it comes to deciding the priority of a rule. For instance, if confidence and support of two rules are same, then the priority of rule r1 is higher than that of rule r2 if the length of antecedent of rule r1 is greater than that of rule r2.

3.3 ACAC Classifier (Associative Classification based on All-Confidence

This classification process is carried out using the All-Confidence metric. It is divided into two phases: the rule generation phase, which is carried out using the Apriori algorithm, and the classification phase, which is based on multiple rules. Typically, the value of an attribute pair is assigned to an item. An item is matched by a data object to the attributes' values. A data item is said to match itemset X

= ai1 ,...,aik , if and only if for (1≤j≤k), the item has value aij in attribute Atij . The total number of objects in the collection that are related to the itemset X is known as the support of X. All confidence is measured as

$$allconf(X) = sup\ (X)\ max\ (sup\ (ai1),\ldots,sup\ (aik) \quad (1)$$

The All-Confidence metric is used to measure the min confidence of the rules derived from the itemset. It can be used to mine the multiple associated itemsets.

3.3.1 Classification Rule Generation

The Apriori algorithm is used to implement the ACAC-RG mining process. It adds the confidence measure to the support-confidence framework to make it more efficient. When a rule passes the all-confidence threshold and passes the support support, it is considered a candidate rule and can be classified.

3.3.2 Multiple Rule based Classification

The ACAC process is carried out by collecting the subset of rules that are related to the new object. It then an associated label to the new item will be assigned, if the rules follow the same class label. If the rules aren't consistent, then ACAC groups them into categories according to the class labels.

3.4. ACCF (Associative Classification Based on Closed Frequent Itemsets)

The first phase of ACCF involves rule generation. It is carried out by using the Charm algorithm[30] to generate all the CFIs from the training dataset. Afterwards, it mines the tidset data of the various rules and locates the meet of the classes' labels. The ACCF algorithm takes into account the meet of the tidsets to determine the Confidence(R) and Support(R)of the CARs (Class Association Rules). It also selects the ones that pass the confidence and support thresholds. This ensures that the set is always composed of all the possible rules.

3.5. ACN (Associative Classifier With Negative Rules)

The ACN method, which is an efficient method for generating

negative rules, is used to improve the accuracy of the classification process. It requires a sufficient number of high-quality rules to ensure that the system is capable of producing accurate outcomes. The mining process used for generating ACN's negative rules is completely different. The main advantage of this method is that it allows to generate a large number of good and strong negative rules, which can be used in place of weak positive ones. This ensures that the system can produce accurate outcomes.

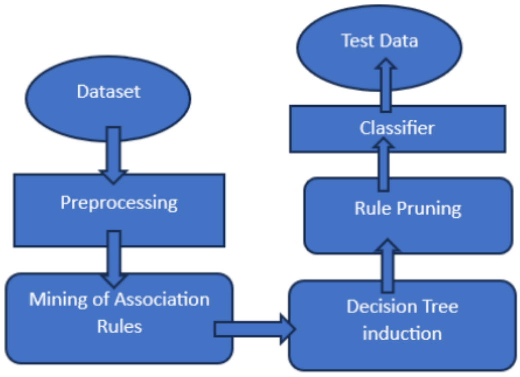

Fig. 1. Working Principle of Associative Classifier

3.6. L3 (Live and Let Live Classifier- Lazy Approach)

The L3 pruning technique is used to discard the rules that only produce erroneous classification outcomes. In the first step, the system checks if the case has been properly classified. It then considers the remaining rules that were not used during the training phase.

4. Experimental Results

4.1 Preprocessing of Data

The SEER Data set is a collection of statistics about cancer incidence in the United States. It is made up of information collected by various cancer registries in the country. These programs collect various details about the patients, such as the type of cancer, its stage, and treatment options. Various attributes such as patient status, tumour site etc are part of the data set that was collected by SEER. The selected attributes for this study have been mentioned in Table 1. Following the implementation the pre-classification procedure, the overall number of records was 754652. Of these, 34813 were categorized as survivors, while 9421 were non-survivable.

4.2 Implementation Details

The paper utilizes the framework known as SPMF (Sequential Pattern Mining Framework) [31], which is a kind of data mining library that focuses on the exploration of patterns in a database. SPMF provide a wide range of Data Mining algorithms specifically in the area of classification , association rule mining and clustering. In the study six association based classification algorithms such as ACAC, ACCF, ACN, L3, CBA and CBA2 are used for the comparative evaluation on the accuracy and space time requirement.

Table 1. Dataset Attributes

Variable Names	No: of distinct values
Race or Ethnicity	29
Primary Site	9
Behavior Code ICD-O-3	1
Grade	9
CS-Extension	35
CS-Lymph Nodes	37
CS-Tumor Size	999
Histology Recode Broad Groupings	31
SEER historic stage A	6
First malignant primary indicator	2
Age Recode	15
Regional Nodes Positive	99
Regional Nodes examined	99
Sequence No	6

As the initial level of this study the complete dataset has been divided into four sample sets where each sample set is arranged randomly with survived and non survived records. From each sample set 80% of records are considered as training set and 20% considered for testing purpose. With a minimum support and confidence level of 0.5 all the six algorithms have been applied on training set as well as testing set. The same have been applied for all four sub datasets (ie. Sample1 as subdataset1, sample 1 and 2 together as subdataset2, sample 1,2 and 3 as subdataset3, sample 1, 2, 3 and 4 as subdataset4) After implementation three main observations have been identified on the basis of accuracy, F1 score and time-space requirement.

4.3 Analysis on accuracy

Considering the accuracy achieved for all four sub datasets for both CBA and CBA2, the study revealed that as the size of the sub dataset increases the accuracy also seem to be increasing. As far as sample1 is concerned for CBA and CBA2, the accuracy reaches

Table 2. Accuracy achieved by CBA and CBA2

Dataset	CBA		CBA2	
	Training	Testing	Training	Testing
Sample 1	0.5013	0.495	0.6816	0.6868
Sample 1 & 2	0.752	0.7364	0.7523	0.7428
Sample 1,2 & 3	0.7988	0.792	0.7988	0.792
Sample1,2,3 & 4	0.8169	0.8165	0.8169	0.8165

only upto 0.5013 and 0.6816 respectively. Bust as the size the of the sub dataset increases it is found that accuracy has reached its maximum of 0.8169 and 0.8165. The comparison is given in the Table 2. which is followed by Fig2.

Accuracy results achieved for the training as well as testing set for all the six algorithms have been given below in Fig 4 and Fig 5. Values are mentioned in Table 4 and Table 5. It is clear from the results that while the training accuracy achieved by ACAC algorithm starts at 0.4886 with sample1 and reach upto 0.7861 for the sample1,2,3&4 dataset(complete dataset), same for ACCF algorithm starts with 0.6351 and reaches upto 0.7861. ACN and L3 algorithms also not showing an advancement in the accuracy achieved for the maximum sized dataset where CBA and CBA3 reaches the maximum accuracy of 0.8165.

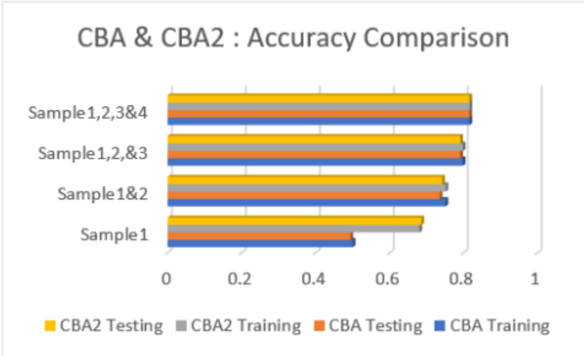

Fig. 2. Accuracy Results for CBA and CBA2

Table 3. Accuracy achieved by CBA and CBA2

Dataset	Naiive Bayes	Random Tree	CBA	CBA2
Sample1	0.703025	0.696656	0.495	0.6868
Sample1&2	0.7592	0.720453	0.7364	0.7428
Sample1,2&3	0.778928	0.7569	0.792	0.792
Sample1,2,3&4	0.796518	0.7773	0.8165	0.8165

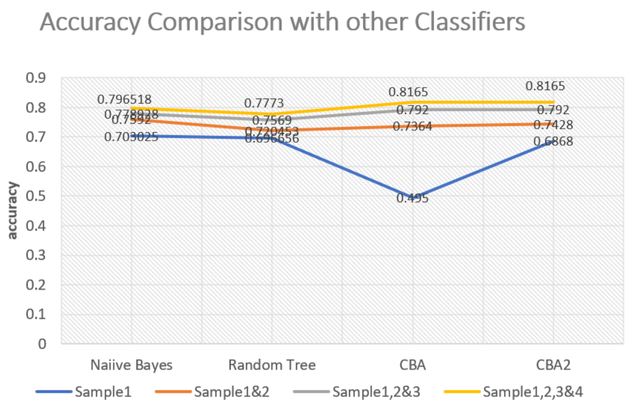

Fig. 3. Accuracy Results for CBA and CBA2compared with traditional classifiers

As far as testing accuracy is concerned, CBA shows a lesser accuracy for sample1 dataset but reaches an accuracy of 0.8165 with maximum sized dataset, whereas CBA2 consistently shows an acceptable accuracy which ranges from 0.6868 to 0.8165.

Table 4. Training accuracy of various models

Dataset	ACAC	ACCF	ACN	L3	CBA	CBA2
Sample1	0.5029	0.6334	0.5029	0.6334	0.5013	0.6816
Sample 1&2	0.6696	0.6551	0.6696	0.6696	0.752	0.7523
Sample 1,2 &3	0.7527	0.7527	0.7527	0.7527	0.7988	0.7988
Sample1,2,3 & 4	0.7873	0.7873	0.7873	0.7873	0.8169	0.8165

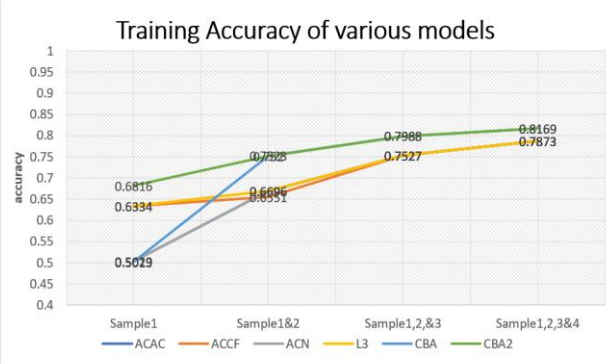

Fig. 4. Training accuracy of various models

Table 5. Testing accuracy of various models

Dataset	ACAC	ACCF	ACN	L3	CBA	CBA2
Sample1	0.4886	0.6351	0.4886	0.6351	0.495	0.6868
Sample 1&2	0.6551	0.6551	0.6551	0.6551	0.7364	0.7428
Sample 1,2 &3	0.7393	0.7393	0.7393	0.7393	0.792	0.792
Sample1,2,3 & 4	0.7861	0.7861	0.7861	0.7861	0.8165	0.8165

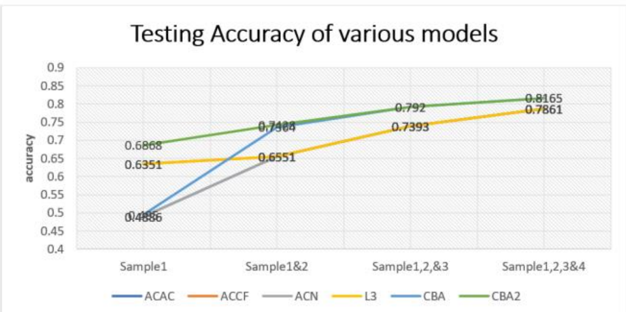

Fig. 5. Testing accuracy of various models

4.4 Analysis on F1 score

Although accuracy provides a more efficient algorithm, it only measures how many times a prediction is correct across the entire dataset. This is because, if the data is class-balanced, the accuracy of the model still remains valid.

The F1 score evaluates a model's predictive skill by highlighting its class-wise excellence instead of its accuracy. This evaluation method has gained widespread application in recent scientific literature. It combines the recall and precision scores of a model to create a single evaluation metric. As the metric value grow closer to 1, that efficient the model will be.

$F1\ Score = 2*(Precision*Recall)\ /Precision+Recall$ (2)

Fig 6 and Fig 7 shows the F1 score achieved for training as well as testing dataset and for CBA and CBA2 the F1 score reaches almost 0.7. F1 Score values are mentioned in Table 6 and Table 7.

Table 6. F1 score attained by given algorithms for training dataset

Dataset	ACAC	ACCF	ACN	L3	CBA	CBA2
Sample1	0.3345	0.6728	0.3345	0.6729	0.3338	0.704
Sample 1&2	0.401	0.401	0.401	0.401	0.7038	0.7007
Sample 1,2 &3	0.42948	0.42948	0.42948	0.42948	0.6978	0.6978
Sample 1,2,3 & 4	0.4404	0.4404	0.4404	0.4404	0.69311	0.69311

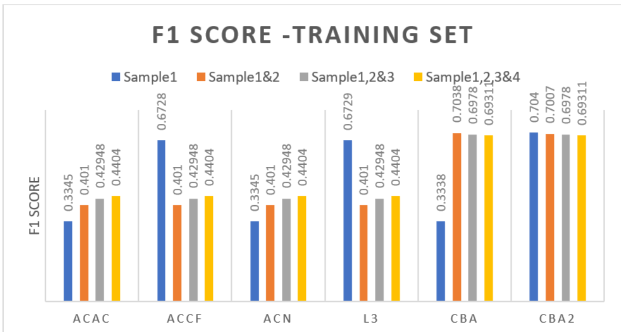

Fig. 6. F1 score achieved by various algorithm for training dataset

Table 7. F1 score attained by given algorithms for testing dataset

Dataset	ACAC	ACCF	ACN	L3	CBA	CBA2
Sample1	0.3282	0.684	0.3282	0.6811	0.3311	0.7069
Sample1 &2	0.3957	0. 3957	0. 3957	0. 3957	0.693	0.6983

Sample1, 2&3	0.42502	0. 42502	0. 42502	0. 42502	0.7011	0. 7011
Sample1, 2,3&4	0.4401	0.4401	0.4401	0.4401	0.69269	0. 69269

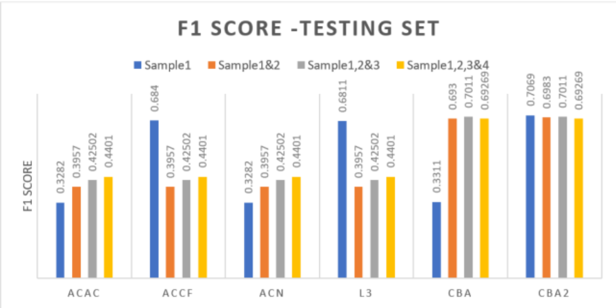

Fig. 7. F1 score achieved by various algorithm for testing dataset

4.5 Analysis on time-space requirement

As far as the training set is concerned , the memory required by the CBA2 algorithm is 21.018 mb only while ACN algorithm demands the highest memory requirement and CBA algorithm demands less time which is only 32 seconds whereas ACAC algorithm demands the highest time requirement. Memory and Time measurements are mentioned in Table 8 and Table 9 and the result comparison is shown in Fig 8 and Fig 9. Considering the testing dataset also it is found that CBA and CBA2 algorithm have taken fairly minimum time and memory which in turn strengthen the fact that CBA and CBA2 are acting more efficiently on increased size datasets.

Table 8. Time-space requirement for the complete training dataset

Model	Memory in mb	Time in ms
ACAC	25.5344	102
ACCF	58.2392	46
ACN	103.585	62
L3	45.9986	47
CBA	301798	32
CBA2	21.018	65

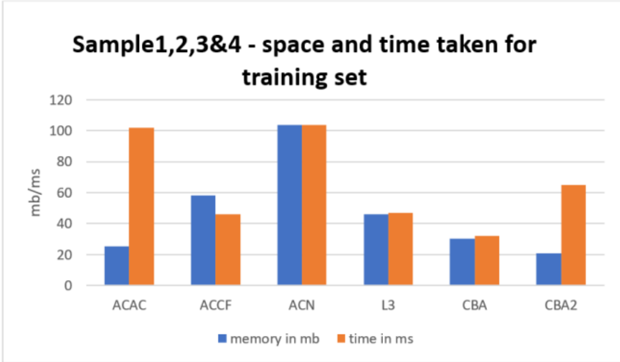

Fig. 8. Time-space requirement for the complete training dataset

5. Conclusion

This study aimed to identify an appropriate associative classifier to classify a breast cancer patient as survivable or non-survivable.

Table 9. Time-space requirement for the complete testing dataset

Model	Memory in mb	Time in ms
ACAC	33.0325	24
ACCF	48.4037	31
ACN	19.3769	27
L3	46.4966	11
CBA	31.1798	7
CBA2	21.839	11

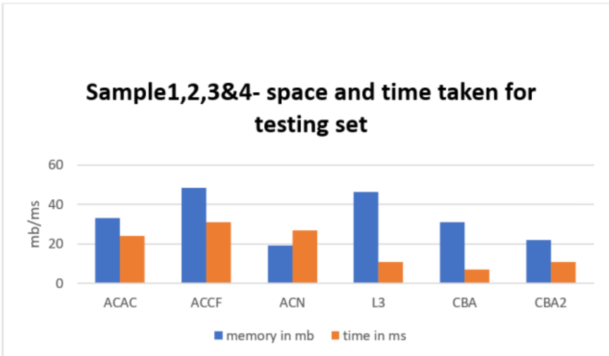

Fig. 9. Time-space requirement for the complete testing dataset

For the study, SEER (2020 November) breast cancer dataset has been used. After applying the pre-classification criteria, the efficiency of CBA and CBA2 algorithms on SEER dataset has been compared with other 4 association based classifiers such as ACAC, ACCF, ACN and L3 . The results revealed that CBA and CBA2 exhibit an accuracy of 81% which is the highest compared to all other four classifiers. As we consider another metric like F1 score concerned with efficiency CBA and CBA 2 show a score value of 0.7 which is closer to 1, compared to other algorithms. It is also found that CBA and CBA2 have a relatively low time and memory consumption which makes them more efficient when it comes to handling large datasets.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Van der Aalst W. “Process Mining: Data Science in Action”.Springer; Berlin/Heidelberg, Germany: 2016.

[2] Romero C., Ventura S. “Educational data science in massive open online courses”. Wires Data Min. Knowl. Discov. 2017;7:1–12. doi: 10.1002/widm.1187.

[3] Bray, F.; Ferlay, J.; Soerjomataram, I. “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. CA Cancer J. Clin. 2018, 68, 394–424.

[4] Blake, C.L.; Merz, C.J. UCI Repository of Machine Learning Databases. 1998. Available online: http://www.ics.uci.edu/$\sim$mlearn/MLRepository.html (accessed on 20 March 2020)

[5] Chen, Hua, et al. "Classification prediction of breast cancer based on machine learning." Computational Intelligence and Neuroscience 2023 (2023).

[6] Bazazeh, D.; Shubair, R. “Comparative study of machine learning algorithms for breast cancer detection and diagnosis”. In Proceedings of the 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, UAE, 6–8 December 2016; pp. 1–4.

[7] Chaurasia, Vikas, and Saurabh Pal. "Data mining techniques: to predict and resolve breast cancer survivability." International Journal of Computer Science and Mobile Computing IJCSMC 3.1 (2014): 10-22.

[8] Delen, Dursun, Glenn Walker, and Amit Kadam. "Predicting breast cancer survivability: a comparison of three data mining methods." Artificial intelligence in medicine 34.2 (2005): 113-127.

[9] Srinivas, K., G. Raghavendra Rao, and A. Govardhan. "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques." 2010 5th International Conference on Computer Science & Education. IEEE, 2010.

[10] Bernal, James Lopez, Steven Cummins, and Antonio Gasparrini. "Interrupted time series regression for the evaluation of public health interventions: a tutorial." International journal of epidemiology 46.1 (2017): 348-355.

[11] Williams, Tom GS, et al. "Risk prediction models for colorectal cancer in people with symptoms: a systematic review." BMC gastroenterology 16 (2016): 1-16.

[12] Oyewola, David, Danladi Hakimi, Kayode Adeboye, and Musa Danjuma Shehu. "Using five machine learning for breast cancer biopsy predictions based on mammographic diagnosis." International Journal of Engineering Technologies IJET 2, no. 4 (2016): 142-145.

[13] Aruna, S., S. P. Rajagopalan, and L. V. Nandakishore. "Knowledge based analysis of various statistical tools in detecting breast cancer." Computer Science & Information Technology 2.2011 (2011): 37-45.

[14] Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H. “Predicting diabetes mellitus with machine learning techniques”. Frontiers in genetics. 2018 Nov 6;9:515.

[15] Asri, Hiba, Hajar Mousannif, Hassan Al Moatassime, and Thomas Noel. "Using machine learning algorithms for breast cancer risk prediction and diagnosis." Procedia Computer Science 83 (2016): 1064-1069.

[16] Wang, H.; Yoon, W.S. “Breast cancer prediction using data mining method”. In Proceedings of the 2015 Industrial and Systems Engineering Research Conference, Nashville, TN, USA, 30 May–2 June 2015.

[17] Mangasarian, O.L.; Street, W.N.; Wolberg, W.H. “Breast cancer diagnosis and prognosis via linear programming”. Oper. Res. 1995, 43, 570–577.

[18] Wolberg, W.H. Wisconsin Breast Cancer Database; University of Wisconsin Hospitals: Madison, WI, USA, 1991.

[19] Nithya, R., and B. Santhi. "Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer." International journal of computer applications 28.6 (2011): 21-25.

[20] Agarap, Abien Fred M. "On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset." Proceedings of the 2nd international conference on machine learning and soft computing. 2018.

[21] Jirı, Filip, and Tomáš Kliegr. "Classification based on associations (cba)-a performance analysis." (2018).

[22] Huang, Zaixiang, et al. "ACAC: Associative Classification based on All-Confidence." 2011 IEEE

International Conference on Granular Computing. IEEE, 2011.

[23] Li, Xueming, Dongxia Qin, and Cun Yu. "ACCF: Associative classification based on closed frequent itemsets." 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery. Vol. 2. IEEE, 2008.

[24] Kundu, Gourab, et al. "ACN: An associative classifier with negative rules." 2008 11th IEEE International Conference on Computational Science and Engineering. IEEE, 2008.

[25] Baralis, Elena, and Paolo Garza. "A lazy approach to pruning classification rules." 2002 IEEE International Conference on Data Mining, 2002. Proceedings.. IEEE, 2002.

[26] Chen Guoqing, Liu Hongyan, Yu Lan, Wei Qiang, Zhang Xing. (2006) “A new approach to classification based on association rule mining.” Decision Support Systems 42(2): 674–689.

[27] Deng Houtao, Runger George, Tuv Eugene, Bannister Wade. (2014) “CBC: an associative classifier with a small number of rules.” Decision Support Systems, 59(1): 163–170.

[28] Ayyat Suzan, Lu Joan, Thabtah, Fadi Abdeljaber. (2014) “Class strength prediction method for associative classification.” in Proc. ARIA the Fourth International Conference on Advances in Information Mining and Management, pp. 5-10, France.

[29] Alcala-Fdez, J., Alcala, R., Herrera, F.: “ A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning”. IEEE Transactions on Fuzzy Systems 19(5), 857–872 (2011)

[30] M.J. Zaki, Ching-Jui Hsiao, “Charm: An efficient Algorithm for closed Association Rules Mining”, Technical Report, TR99-10.Computer Science Dept. Rensselaer Polytechnic Institute, 1999.

[31] Fournier-Viger, Philippe, et al. "Spmf: a java open-source pattern mining library." J. Mach. Learn. Res. 15.1 (2014): 3389-3393.

[32] Abubacker, Nirase Fathima, et al. "Breast cancer detection by using associative classifier with rule refinement method based on relevance feedback." Neural Computing and Applications 34.19 (2022): 16897-16910.

[33] Behar, Nishant, and Manish Shrivastava. "A Novel Model for Breast Cancer Detection and Classification." Engineering, Technology & Applied Science Research 12.6 (2022): 9496-9502.

[34] Wang, Kung-Jeng, Bunjira Makond, and Kung-Min Wang. "An improved survivability prognosis of breast cancer by using sampling and feature selection technique to solve imbalanced patient classification data." BMC medical informatics and decision making 13.1 (2013): 1-14.

[35] Ponnuraja, Chinnaiyan, et al. "Decision tree classification and model evaluation for breast cancer survivability: A data mining approach." Biomedical and Pharmacology Journal 10.1 (2017): 281-289.