An SVM Classifier With Autoencoder Feature Extractor for Breast Cancer Diagnosis

Murilo Venturin
*PostGraduate Program in Biomedical
Engineering (PPGEB)
University of Brasília
Brasilia, Brazil*
ORCID: 0000-0002-4124-826X

Luciana A. Fernandes
*PostGraduate Program in Biomedical
Engineering (PPGEB)
University of Brasília
Brasilia, Brazil*
ORCID: 0000-0003-1291-1912

Rebeca H. M. Oliveira
*Biomedical Engineering Department
(BME)
Johns Hopkins University
Maryland, United States*
ORCID: 0000-0001-7239-3462

Ronei D. Fonseca Campos
*PostGraduate Program in
Mechatronics Engineering (PPMEC)
University of Brasília
Brasilia, Brazil*
ORCID: 0000-0001-8835-9031

Suélia S.R.F. Rosa
*PostGraduate Program in Biomedical
Engineering (PPGEB)
University of Brasília
Brasilia, Brazil*
ORCID: 0000-0002-1247-9050

Abstract—**In this paper, we present the development and training of an efficient algorithm based on Support Vector Machine (SVM) with dimensionality reduction for auto encoder for application on the process of Breast Cancer classification. We use Wisconsin Diagnostic Breast Cancer (WDBC) as our original data set. Detection and characterization of benign and malignant for the diagnosis of breast cancer is important for early diagnosis. Five autoencoder models based on fully connected neural networks with different data compression rates were developed. After being trained, the characteristic vector of each model was extracted, so an SVM is used to classify the type of tumor from that vector. to validate the models, 10-fold cross validation was used. From the original data set, we used the trained encoder to extract vectors of size three, five, eight, ten and sixteen. The results obtained are compared with other results found in the literature and showed a better performance, reaching an accuracy of 99.12%. Finally, the experimental results illustrate that the proposed algorithm has the ability to compress the original data without losing information relevant to the classification, presenting a better performance than the traditional SVM for the classification of breast cancer.**

Keywords — Autoencoder, Breast Cancer, Deep Learning, Dimensionality Reduction, Machine Learning, SVM

I. Introduction

Breast Cancer (BC) is the second type of cancer with a greater number of cases reported worldwide, with about 25% of women suffering from it [1]. About 11.6% of all cancer cases are classified as BC [2] and about 6.60% of these cases have led to death [1]. Therefore, novel technologies for improving diagnosis precision of BC are of interest to medicine. [3] [4].

Despite mammography currently being the primary medical intervention for diagnosing Breast Cancer, some issues have been reported on the procedure. In Zheng (2014) and Radiya (2017), one of the greatest problems reported is the divergence of diagnosis between radiologists, leading to imprecise exam results [4], [5]. Incorrect diagnoses as well as divergence in health professionals opinions may lead to harmful impacts for the patient, such as the absence of treatment for an active tumor or improper treatment [5], [6].

Artificial Intelligence (AI) has been one of the strategies used for diagnosing and classifying breast tumors as malignant and benign (binary classification) [3], [4], [7].

Benign cases are those with a reported abnormal growth of the same type of cell, characterized for being a slower process than the Malignant tumor. Malignant cases, contrarily, refer to tumors caused by an abnormal growth that extends to other cell types and tissues of the body, in a faster process that deforms their nuclei [1], [8]. Processing tissue's characteristics, like texture, area, and softness, is essential for classifying BC as benign or malignant [3].

Among the Machine-learning algorithms most currently used for diagnosing and classifying Breast Cancer are Support Vector Machine (SVM), Decision Tree, Gradient Boosting Classifier and Stochastic Gradient Descent (SGD) classifier. Data mining techniques and Extreme Learning Machine (ELM) have also been used for tumor classification [8], [9]. Mert et al. (2011) presented results on using SVM for diagnosing BC. Their work shows that SVM with quadratic kernel presented an accuracy of 94.40% [10].

The authors [7] proposed a paper on Breast Cancer identification utilizing different algorithms to increase data accuracy. They have utilized the K Nearest Neighbor algorithm and decision tree to identify whether the predicted cancer is of the malignant or benign type. They used 32 attributes and 569 data from the Wisconsin Breast Cancer data set. The classification of breast cancer they performed had 12 inputs, in which they calculate the mean and standard deviation. The classification was performed as formulas of the algorithms used. When comparing two algorithms, they obtain an accuracy of 93.85% in the nearest K-neighbor algorithm and 95.61% in the decision tree algorithm.

For classification purposes, SVM has also presented the greatest accuracy when compared to methods such as Random Forest and Naïve Bayes [11]. SVM has also been compared to Decision Tree for classification purposes, in an analysis based on the Wisconsin Breast Cancer dataset, presenting a 97.13% accuracy [12]. This contrasts with the findings of Sathiyanarayanan *et al*. (2019), who obtained a 99% accuracy on training with classification by using the Decision Tree algorithm [13] and Maheshwar (2019), who also found better accuracy with Decision Tree classifiers [14].

For improving accuracy, learning algorithms such as Gradient Boosting Classifier have been used, combining predictive methods [15]. In Guzel (2018), different boosting classifiers are compared for BC classification, with XGBoost presenting the best performance [16]. Gradient Boosting has

also been used as a classifier for analyzing the Wisconsin Breast Cancer dataset accuracy improvement through oversampling [17].

SGD algorithms have been assessed as classifiers for Breast Cancer in [18], [19]. The results obtained by Mittal (2015) indicate better performance for SGD hybridized with other classifiers [19]. The method has also been used together with deep learning algorithms for classifying BC [18].

The study performed by Toprak (2018) was based on achieving classification of BC by using Extreme Learning Machine, comparing results to other techniques (Naive Bayes, SVM and Artificial Neural Network). For the analysis, he also used the Breast Cancer Wisconsin (Diagnostic) data set. The ELM method presented the highest performance between methods, with a 98.99% accuracy. Additionally, the method presented the shortest test time (0.0052 seconds against 0.06 seconds for the SVM and 0.04 seconds for Naive Bayes).

Performing an accurate and early diagnosis as well as proper treatment for BC are important factors for reducing the number of deaths caused by the disease. There are various Machine Learning Techniques used for classifying the type of tumor, and it is not our intent to reject the methods previously proposed by other authors. Instead, we aim to contribute to the current methods by presenting a novel classification technique for Breast Cancer that comprises the union of an autoencoder for dimensionality reduction with the Support Vector Machine technique. This approach, as far as we know, has not been used for analyzing this particular scenario, and might provides us an increased accuracy for data classification.

II. METHODOLOGY

A. Data Description

The data used was extracted from the Diagnostic Wisconsin Breast Cancer Database. Following, we provide a brief description of this data. Tissue characteristics were obtained by means of a digital image of a fine-needle aspiration exam of breast mass. A total of 569 patients (212 malignant and 357 benign) were analyzed: 212 with breast cancer and 357 with fibrocystic breast masses, as described by [20].

Ten original characteristics of each tumor were quantified, which are:

- Radius (mean of distances from center to points on the perimeter);

- Texture (standard deviation of gray-scale values);

- Perimeter;

- Area;

- Smoothness (local variation in radius lengths);

- Compactness (perimeter2/area − 1.0);

- Concavity (severity of concave portions of the contour);

- Concave points (number of concave portions of the contour);

- Symmetry;

- Fractal dimension ("coastline approximation" - 1).

From the original data provided by [20], the mean, standard error and the largest mean of the three largest values of all the original features were computed for each image. The dataset with 30 characteristics, in addition to the label (benign and malignant), was provided by the Center for Machine Learning and Intelligent Systems [21].

B. Data Pre-Processing

Using the stratified K-fold cross-validation technique with k=10, we divided the dataset into 10 validation and training groups, each group contained 57 samples for validation and 512 samples for training, with the exception of one group that was left with 56 samples for validation and 513 for training. A visual representation of the k-fold technique can be seen in Fig. 1. From these groups, we trained and validated each model mentioned in this article 10 times, once for each group. We obtained the performance metrics of the models using a simple measurement of arithmetic average of the 10 metrics generated by the validation group in their respective trained model.
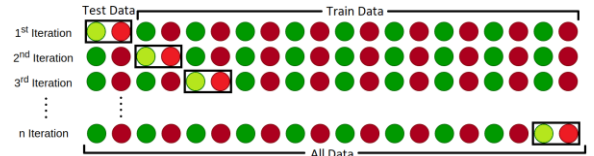

Fig. 1. Representation of the K-fold cross-validation method.

We submitted the 30 characteristics of each sample to an scale transformation, so that every datum of the train dataset was in the interval [0,1]. For that, we subtracted from each value the minimum value present on the train dataset. We divided the results by the difference between the maximum and minimum values on the training dataset.

This process is shown in equation 1, where $X_{re-scaled}$ represents the new value, X_i the original value, $min(X)$ the lowest present value in the training dataset and $max(X)$ the highest present value also in the training dataset. We substitute each original value by the final result obtained.

$$X_{re-scaled} = \frac{X_i - min(X)}{max(X) - min(X)} \qquad (1)$$

After re-scaling data by applying Eq. 1 on the training group, we repeated the same transformation to the validation group. We defined $min(X)$ and $max(X)$ values previously, based on the training group. We performed this process in all training groups and their validation correspondents. The scale transformation preserves the original distribution without a significant alteration to the information brought by each value. It also has as advantage not reducing the importance of discrepant data (outliers).

C. Dimensionality Reduction With Deep Learning Model

We created five different autoencoder models to reduce the dimensionality of the data set and improve the performance of the [22] classifier model. The encoding and decoding parts of the autoencoders were built with fully connected neural networks.

First, the input data passes through the encoder, which reduces its dimensionality. The developed models reduce the input dimensionality from 30 to 3 (autoencoder-3D), 5 (autoencoder-5D), 8 (autoencoder-8D), 10 (autoencoder10D) and 16 (autoencoder-16D).

The decoder has a mirrored encoder architecture and is responsible for reconstructing the 30 input characteristics to obtain an output as close as possible to the original input.

Our autoencoders have seven fully connected layers. The first (input layer) consists of 30 neurons, the same amount as the number of resources in our input data and an activation function of the Rectified Linear Unit (ReLU). The second layer has 16 neurons, while the third layer has 8 neurons in the autoencoder-3D, autoencoder-5D and autoencoder8D models, and 16 neurons in the autoencoder-10D and autoencoder-16D models. Both use ReLU as the activation function.

The fourth layer is the smallest of our auto encoder, representing the encoder output. For this layer, we use a sigmoid activation function, so that the data is restricted to values between 0 and 1. The number of neurons in this layer in each model is the number corresponding to the encoder output dimensionality.

The fifth layer has 8 neurons in the autoencoder-3D, autoencoder-5D and autoencoder-8D models and 16 neurons in the autoencoder-10D and autoencoder-16D models, all with the ReLU activation function. The sixth layer has 16 neurons in all models also with the ReLU activation function.

The output layer (seventh layer) has 30 neurons, the same amount as in the input layer. The output must be the same as the input. The linear activation function is used in this layer.

Table 1 provides a description of the architecture of each autoencoder model, while Fig. 2 provides a graphical representation of the autoencoder-3D architecture and its layers, the graphic representation of the other models was considered unnecessary due to similarity between models.

TABLE I. ARCHITECTURE OF THE PROPOSED AUTOENCODER MODELS, WITH SEVEN FULLY CONNECTED LAYERS EACH AND DIFFERENT COMPRESSION RATES OF THE INPUT DATA. THE FOURTH LAYER OF EACH MODEL REPRESENTS THE ENCODER OUTPUT, FROM WHICH THE CHARACTERISTIC VECTOR IS EXTRACTED.

Model	Layer						
	1	2	3	4	5	6	7
Autoencoder-3D	30	16	8	3	8	16	30
Autoencoder-5D	30	16	8	5	8	16	30
Autoencoder-8D	30	16	8	8	8	16	30
Autoencoder-10D	30	16	16	10	16	16	30
Autoencoder-16D	30	16	16	16	16	16	30

As in Artificial Neural Networks (ANN), autoencoders are trained with backpropagation. We used the mean square error as the loss function, which is calculated by:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2, \qquad (2)$$

where N is the amount of data, fi is the value returned by the model (decoder output) and y_i is the real value of the datum. The data previously divided into 10 training and validation groups and re-scaled were used for training each autoencoder model 10 times. he models were trained by 2000 epochs with batches of 513 samples, hence, at each epoch the model weight was recalculated only once. We used Adam optimization algorithm [23], which finds an adaptable optimization rate for each parameter. After being trained, the encoder for each model was used to extract the characteristic vector with dimensions of 3, 5, 8, 10 and 16 from the entire

dataset, maintaining the training and validation division. With these vectors, we can train a machine learning classification model.
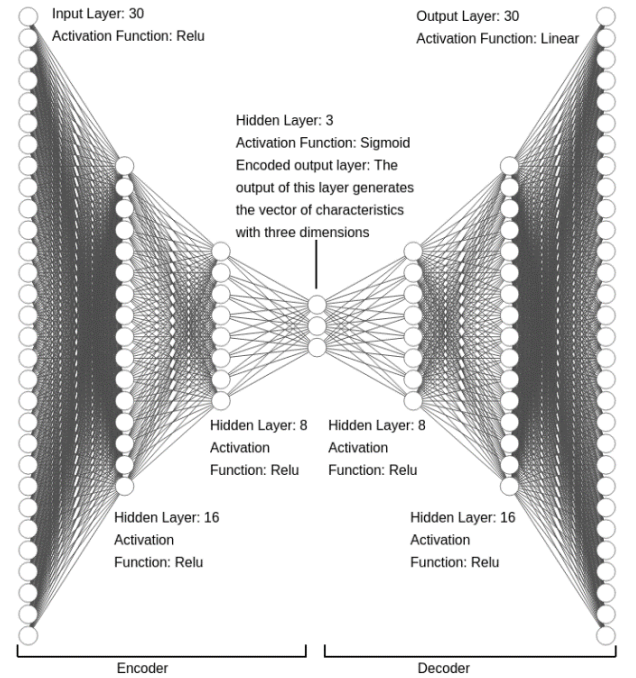

Fig. 2. Autoencoder-3D model architecture fully connected to the central layer, which is the encoder's output, with the dimension equal to 3.

D. Classification Model

For classifying between malignant and benign BC, we recurred to the Machine Learning model of Support Vector Machine (SVM) [24]. During training, a SVM uses supervised learning to define the ideal hyperplane for separating data.

We build the SVM through a radial basis function kernel (RBF kernel) with the gamma coefficient set to 'scale'. The regularization parameter was set to 1.0 and tolerance for stopping criteria was set to 0.001.

The SVM model was trained 10 times for each autoencoder model, using the latent vector extracted by the encoder and respecting the training and validation groups defined initially.

X refers to the training vector, while Y refers to the label related to X. The label Y was set as the diagnosis column of the dataset. The values in this column were set either to M (malignant) or B (benignant) which were altered respectively to 0 and 1.

After training, each SVM model was used to predict the class (0 = malignant and 1 = benign) of the validation group. The comparison between the predicted values and the actual values generated the performance metrics of the method as a whole.

Fig. 3 represents all the steps of the proposed methodology. In order to classify breast cancer between malignant and benign using the dataset provided by the Center for Machine Learning and Intelligent Systems [21]. The proposed method uses a hybrid model of autoencoder and SVM.
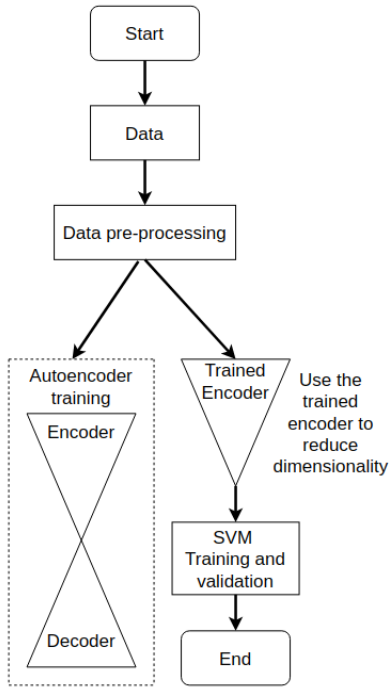
Fig. 3. The figure represents the proposed methodology. For each autoencoder, the models were trained and validated ten times.

E. Performance Evaluation

To evaluate the performance of the classifying model, we used criteria of accuracy, precision, recall and F1-score. As mentioned earlier, autoencoder-SVM was trained and validated using k-fold stratified cross-validation with 10 iterations, so the metrics for each model is the simple arithmetic mean of the metrics for each fold. To perform this assessment only the results of the test data were considered. The accuracy of each training was calculated by the equation 3, where T P: True Positive, T N: True Negative, F P: False Positive, and F N: False Negative.

$$Acuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (3)$$

Precision was calculated using the equation 4.

$$Precision = \frac{TP}{TP+FP} \qquad (4)$$

Recall was calculated using the equation 5.

$$Recall = \frac{TP}{TP+FN} \qquad (5)$$

F1-score was calculated using the equation 6.

$$F1_Score = 2 * \frac{Precision*Recall}{Precision+Recall} \qquad (6)$$

III. RESULTS

As shown in Fig. 3, we trained five autoencoders, after performing pre-processing on the dataset used in our experiment, to reduce the dimensionality of the original data and thereby removing redundancy from the original features [21]. In addition, we used the data compressed by the trained autoencoder to train a Support Vector Machine (SVM) [25] in order to classify the breast tumors evaluated between malignant and benign.

We performed a stratified k-fold cross-validation [26] with k = 10 to validate the classification model adopted, that is, we trained the autoencoders, followed by the SVM 10 times (one training for each fold) and then their results were averaged.

A. Dimensionality Reduction With Deep Learning Model

According to Table 1 and Fig. 2 we reduced the original data of 30 features to a three, five, eight, ten and sixteen-dimensional vector. The autoencoders presented the average loss of the test groups of the 10 folders between 0.0079 for the autoencoder3D and 0.0020 for the autoencoder-16D. The difference between the loss of autoencoders is not so great even with a big difference in the reduction of dimensionality. Fig. 4 Represents the average loss values of all autoencoder models in the training and test groups. Since the loss is calculated by the mean square error, the value obtained represents the error between the input and output data, it demonstrates that there was no overfitting in any of the models and that there is also no big difference in the loss values between the models.
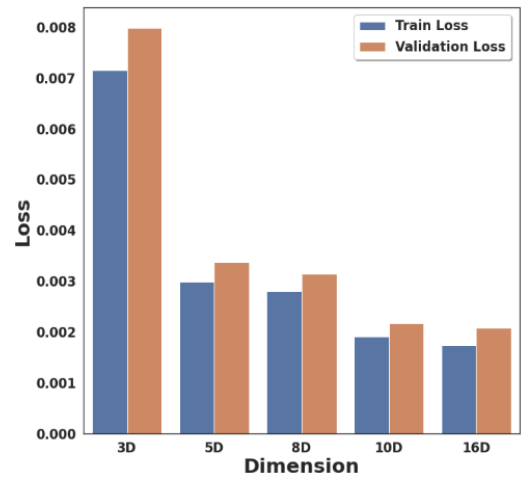

Fig. 4. Result of training and testing of autoencoders, evaluating the loss during 2000 epochs. The loss value refers to the simple arithmetic mean of the 10-fold cross-validation for each model.

By reducing dimensionality of the hidden layer, the autoencoder learned only important information from the dataset. Fig. 4 shows that, the autoencoder was able to identify and obtain the main informations without significant losses. Even with a dimensionality reduction reaching the order of 10 times in the case of the autoencoder-3D, the model was able to compress and decompress the data with a small reconstruction error, demonstrating that the autoencoder was able to compress the data well.

After using the encoder to extract the latent vector (compressed layer of the autoencoder) and reducing data dimension to a three, five, eight, ten and sixteen-dimensional array, we were able to observe visually the separation between Malignant and Benign data for autoencoder-3D (Fig. 5) indicates that the autoencoder had an appropriate performance on dimensionality reduction.
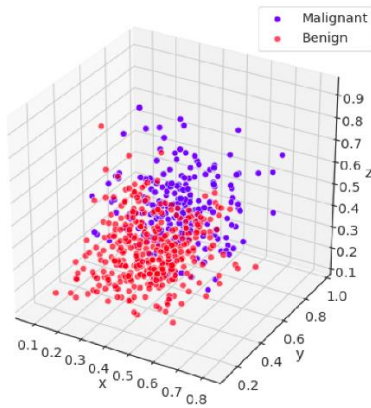
Fig. 5. 3D distribution of the 569 examples after dimensionality reduction by the encoder of autoencoder-3D. for this representation, only the test data for each fold of the 10-fold cross-validation was considered.

B. Classification With Machine Learning Model

As mentioned earlier, an SVM machine learning model was trained to classify the type of tumor from the encoder output data with different dimensions. The metrics obtained in the 10 folds with cross validation can be observed graphically in Fig. 6.
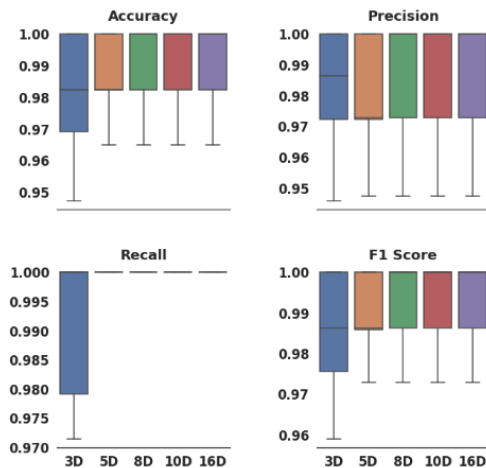

Fig. 6. Box plot of metrics accuracy, precision, recall and f1-score for each model. For this representation, the metrics of the 10-fold cross-validation test groups were considered.

According to Fig. 6 it is possible to observe that even with the reduction of the dimensionality of the data in the order of 10 times the presented technique demonstrated high performance. The metrics demonstrate that the model is capable of classifying tumors with good quality and high stability. The final metrics for each model were obtained using the simple arithmetic mean of the test group results for each fold and can be seen in Table 2.

TABLE II. COMPARISON BETWEEN THE METRICS OF THE AUTOENCODER-SVM MODELS DEVELOPED WITH DIFFERENT DIMENSIONS. THE SIMPLE ARITHMETIC MEAN OF THE TEST DATA FOR THE 10-FOLD CROSS-VALIDATION OF EACH METRIC OF EACH MODEL WAS CONSIDERED.

Autoencoder Model	Metrics			
	Accuracy (%)	Precision	Recall	F1-Score
3D-SVM	98.24	98.10	99.16	98.61
5D-SVM	98.76	98.10	1.0	99.04
8D-SVM	99.12	98.65	1.0	99.31
10D-SVM	99.12	98.65	1.0	99.31
16D-SVM	99.12	98.65	1.0	99.31

It can be seen in Fig. 6 that the model presented a recall very close to one for the autoencoder-3D and equal to one in the other models, which demonstrates that the amount of false negatives was very close to zero or equal to zero. The precision varies between 98.10 and 98.65, indicating that there were not many false positives, but in general the models presented more false positives than false negatives.

It was expected to obtain a recall greater than the precision since the dataset is unbalanced, with 62.74% of the data being from benign tumors and only 37.26% from malignant tumors. So naturally the model had more data on benign tumors to train.

It is possible to observe that the metrics improve with less compression of the original data by the autoencoder, but from the autoencoder-8D the metrics are the same for the other models (autoencoder-10D and autoencoder-16D). This means that for the proposed model the ideal compression is for 8 dimensions, less than that important data is lost and affects the classification and above that redundant and unnecessary data is considered.

The 10-fold cross-validation strategy we use, if compared to alternatives presented in the literature, such as the works of [4] and [24], indicates a superior accuracy and reduction of spatial dimension. Such comparison in presented at Table 3.

The other metrics present in this work could not be compared to other works, since we were not able to find them presented in this particular scenario by other authors.

TABLE III. COMPARISON OF PERFORMANCE AND DIMENSIONALITY REDUCTION BETWEEN THE PROPOSED MODELS AND THE MODELS IN THE LITERATURE.

Approach	Spatial dimension	Accuracy (%)
Autoenconder-SVM	3	98.24
Autoenconder-SVM	5	98.76
Autoenconder-SVM	8	99.12
Autoenconder-SVM	10	99.12
Autoenconder-SVM	16	99.12
K-SVM	6	97.38
PSO-SVM	17	97.37
GA-SVM	18	97.19
ACO-SVM	15	95.96

Autoencoder-SVM transforms the original data into a new format, reducing the size of the SVM input data. This contributes to a reduction in computational cost and inference time. Based on the data presented, we see that the proposed methodology provided a better performance in the classification of the tumor (malignant and benign), with greater accuracy than the other models presented by [4] and [24] even in the most extreme case where the data compression rate was 10 times, as shown in Table 3.

In the case of median compression where the original data were reduced from 30 to 8, 10 and 16 dimensions, the model showed an accuracy of 99.12% and an F1-score of 99.31. Which demonstrates the high data compression capacity of the autoencoder and the high performance of the hybrid Autoencoder-SVM model.

The proposed technique presents a differentiated way of selecting characteristics to classify breast cancer tumors. We collect the most significant information from the data by means of an autoencoder [21] in order to facilitate in the classification process, which continues later using the support vector machine (SVM) technique.

Autoencoder has a very important role, because, in addition to reducing the dimensionality of the input data of the SVM classifier, it makes it possible to discard redundant information from the original data, making the data samples used to train the SVM more separable. With the compression caused by the autoencoder (which presents a deep learning architecture in this work), it was possible to build a model for classifying breast cancer tumors with an accuracy result comparable to results found in the literature.

IV. CONCLUSION

This paper presents a hybrid model of an autoencoder and SVM technique for classifying Breast Cancer. This method can be used to help health professionals to classify malignant and benign cells, generating a faster and more accurate diagnosis. We compared the resulting model with other traditional methods of dimensionality reduction and data classification applied to the same issue. Our model presented better accuracy for classification purposes.

The dimension reduction using the autoencoder together with SVM presented better performance if compared to the strategies presented in Table 3. This was due to the feature selection performed by the autoencoder, which excluded unnecessary information that is irrelevant for classification purposes.

The study is limited by the fact that essential rules/parameters for classifying BC as benign or malignant have not been defined yet by health professionals. The definition of a set of standard parameters would allow for a reduction of inference time, allowing us to develop a better model for data extraction and tumor classification.

REFERENCES

[1] M. Toğaçar e B. Ergen, "Deep Learning Approach for Classification of Breast Cancer," em *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 2018.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre e A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians,* vol. 68, p. 394–424, 2018.

[3] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani e M. Faisal Nagi, "Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms," *Journal of Healthcare Engineering,* vol. 2019, 2019.

[4] B. Zheng, S. W. Yoon e S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," *Expert Systems with Applications,* vol. 41, p. 1476–1482, 2014.

[5] E. Radiya-Dixit, D. Zhu e A. H. Beck, "Automated classification of benign and malignant proliferative breast lesions," *Scientific reports,* vol. 7, p. 1–8, 2017.

[6] M. E. Brennan, R. M. Turner, S. Ciatto, M. L. Marinovich, J. R. French, P. Macaskill e N. Houssami, "Ductal Carcinoma in Situ at Core-Needle Biopsy: Meta-Analysis of Underestimation and Predictors of Invasive Breast Cancer," *Radiology,* vol. 260, pp. 119-128, 2011.

[7] P. Sathiyanarayanan, S. Pavithra., M. S. A. I. SARANYA. e M. Makeswari., "Identification of Breast Cancer Using The Decision Tree Algorithm," em *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019.

[8] V. Chaurasia, S. Pal e B. B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," *Journal of Algorithms and Computational Technology,* vol. 12, p. 119–126, 2018.

[9] A. Toprak, "Extreme Learning Machine (ELM)-Based Classification of Benign and Malignant Cells in Breast Cancer," *Medical Science Monitor,* vol. 24, p. 6537–6543, September 2018.

[10] A. Mert, N. Kilic e A. Akan, "Breast cancer classification by using support vector machines with reduced dimension," em *Proceedings ELMAR-2011*, 2011.

[11] Y. Khourdifi e M. Bahaj, "Feature Selection with Fast Correlation-Based Filter for Breast Cancer Prediction and Classification Using Machine Learning Algorithms," em *2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, 2018.

[12] A. Bharat, N. Pooja e R. A. Reddy, "Using Machine Learning algorithms for breast cancer risk prediction and diagnosis," em *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, 2018.

[13] P. Sathiyanarayanan, S. Pavithra, M. S. A. I. SARANYA e M. Makeswari, "Identification of Breast Cancer Using The Decision Tree Algorithm," em *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019.

[14] Maheshwar and G. Kumar, "Breast Cancer Detection Using Decision Tree, Naïve Bayes, KNN and SVM Classifiers: A Comparative Study," em *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2019.

[15] F. Y. A'la, A. E. Permanasari e N. A. Setiawan, "A Comparative Analysis of Tree-based Machine Learning Algorithms for Breast Cancer Detection," em *2019 12th International Conference on Information Communication Technology and System (ICTS)*, 2019.

[16] B. E. K. Güzel e D. Önder, "Performance comparision of boosting classifiers on breast termography images," em *2018 26th Signal Processing and Communications Applications Conference (SIU)*, 2018.

[17] N. Cahyana, S. Khomsah e A. S. Aribowo, "Improving Imbalanced Dataset Classification Using Oversampling and Gradient Boosting," em *2019 5th International Conference on Science in Information Technology (ICSITech)*, 2019.

[18] N. Darapureddy, N. Karatapu e T. K. Battula, "Implementation of optimization algorithms on Wisconsin Breast cancer dataset using deep neural network," em *2019 4th International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT)*, 2019.

[19] D. Mittal, D. Gaurav e S. S. Roy, "An effective hybridized classifier for breast cancer diagnosis," em *2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, 2015.

[20] W. H. Wolberg, W. N. Street, D. M. Heisey e O. L. Mangasarian, "Computer-derived nuclear features distinguish malignant from benign breast cytology," *Human Pathology,* vol. 26, p. 792–796, 1995.

[21] D. Dua e C. Graff, *UCI Machine Learning Repository,* 2017.

[22] J. Masci, U. Meier, D. Cireşan e J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," em *International conference on artificial neural networks*, 2011.

[23] D. P. Kingma e J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[24] Y. Prasad, K. K. Biswas e C. K. Jain, "SVM classifier based feature selection using GA, ACO and PSO for siRNA design," em *International conference in swarm intelligence*, 2010.

[25] J. Shawe-Taylor e N. Cristianini, Support vector machines, vol. 2, Cambridge University Press Cambridge, 2000.

[26] N. A. Diamantidis, D. Karlis e E. A. Giakoumakis, "Unsupervised stratification of cross-validation for accuracy estimation," *Artificial Intelligence,* vol. 116, p. 1–16, 2000.