Knowledge extraction about patients surviving breast cancer treatment through an autonomous fuzzy neural network

1st Paulo Vitor de Campos Souza
Depart. Knowledge-Based Math.l Systems
Johannes Kepler University Linz
Linz, Austria
0000-0002-7343-5844

2nd Yu-Kai Wang
Centre for Artificial Intelligence
University of Technology Sydney
Sydney, Australia
0000-0001-8390-2664

3rd Edwin Lughofer
Depart. Knowledge-Based Math.l Systems
Johannes Kepler University Linz
Linz, Austria
0000-0003-1560-5136

***Abstract*—Cancer treatment is extremely aggressive and, in addition to causing considerable discomfort, can lead to death. Therefore, identifying aspects related to treatment assertiveness may be efficient for reducing the mortality rate of cancer patients. This paper seeks to identify the prognosis of cancer treatment survival through hybrid techniques based on the autonomous fuzzification process and artificial neural networks. The public dataset on cancer mortality is the source for conducting treatment assertiveness rating tests. The hybrid model had its results compared to other models present in the pattern classification literature with superior accuracy and identification of people likely to survive treatment (90.46%), and the fuzzy rules obtained with the execution of the model corroborate the high assertiveness of the model, even surpassing state of the art for the theme.**

***Index Terms*—Fuzzy Neural Network, Breast Cancer, Cancer Mortality, Lymph Node Rate.**

I. Introduction

Breast cancer is a significant cause of female mortality worldwide, mainly due to the aggressiveness of the disease or late diagnosis [1]. Treatments for this type of the disease are unpleasant, complicated, and do not always guarantee that cancer will not spread to other parts of the body. Thus, several researchers seek to understand the relationships present like this complex problem to understand them and seek efficient solutions to increase the chance of treatment success and, consequently, the quality of life of women undergoing it [2]. Research on the effectiveness of treatments in women with breast cancer is growing at the same rate as cases of malignancy worldwide. Consequently, there are medical studies that are carried out jointly by medical teams and people in the statistical or computational areas to reproduce intelligent models capable of streamlining and modeling the efficiency of breast cancer treatments. In science, investigations have been conducted to judge psychosocial aspects and their connection to treatment assertiveness, post-treatment effects, relationships, and impacts on sleep disorders during cancer treatment [3]. Some studies assess the cognitive aspects of people who have survived treatment [4]. Another line of research recently addresses aspects of comorbidities, age, and period of diagnosis-influence treatment in the identification of early cancer [5] and, finally, investigates the impacts of hormone replacement for breast cancer cell growth in menopause.

Investigations on the effectiveness of breast cancer treatments can address various social, economic, or women's health aspects. However, it is vital to highlight the role of research based on data collected from specific patient groups. The dataset used in the research of Teng et al. [6], evaluated the breast cancer cases of 5,279 women with infiltrated duct and lobular carcinoma. They were diagnosed from 2006 to 2010, and their data were obtained from the NCI SEER Cancer Registry. The original study sought to use prognostic modeling based on Bayesian inference. The main objective of the research was to estimate the impact of the lymph node ratio (LNR) on the survival of women who underwent breast cancer treatments. For data evaluation purposes, was applied the C-statistic with six methods [1], with high AUC indices (AUC values > 0.7). However, despite the useful results of the Bayesian approach in predicting breast cancer survival, they were poorly interpreted, especially regarding the existing correlations with the other collected dimensions of the problem, such as age, for example. The studies conducted by the authors evaluated three LNR factors with the dimensions of the problem: estimated, measured, and without LNRs. Notwithstanding, this kind of assessment makes the interpretability of the problem confusing. It is not possible to establish dynamic relationships between the problem and the LNR features.

On the other hand, the analysis with estimated factors in this paper aims to evaluate the prognostic value of the LNR through a hybrid model capable of extracting knowledge from the dataset. Thus, the model is expected to act with the full dataset provided by Teng et al. [6]. The original proposal used a Bayesian approach to these assessments making a relationship between the predictor variables and

[1]-Classic Cox proportional hazards regression model.- Modification of Therneau's Coxph function, the Andersen-Gill model. - Random Survival Forest. - Multi-Task Learning Model for Survival Analysis. - Bayesian estimation regression model and the Bayesian approach, from the R-Cran package [6].

different levels of LNR. On the other hand, this paper seeks to evaluate the relationships present in the dataset through a hybrid model that combines fuzzification techniques to identify existing relationships and solve them through neural network training. Thus, the model is capable of generating IF-THEN rules capable of transmitting the existing relationships in the studied database, allowing the results to be understandable and without using Bayesian approaches to estimate parameters. Fuzzy neural networks (FNN) have an efficient performance in the extraction of characteristics and correlations in databases linked to health treatments. Recently intelligent hybrid models have been used in predicting assertiveness motor problems in children [7] and fetal health monitoring [8]. Other models worked with contexts of electroencephalography (EEG) [9] and detecting autism in children [10], and adults [11], and adolescents [12]. Finally, it solved problems in predicting breast cancer, although using other evaluation criteria such as resistin, glucose, age, and BMI proposed by Silva Araujo et al. [13]. Therefore, such an approach can act assertively in the identification of breast cancer Survival Prognosis through Lymph Node Ratio Estimation analysis.

The main contribution of this paper is to bring other evaluation dimensions to the lymph node ratio problem in the survival involvement of women with breast cancer using all features of the problem. Through the fuzzy rules to be obtained, it is hoped that the relationships obtained can clarify the existing correlations between the collected data and their impacts on female mortality in such complex treatments. Another highlight of this work is to seek an intelligent approach that can address the problem completely, thus seeking more positive ways of treatment approaches and extend the survival of people who are affected with breast cancer. It should also be noted that the fuzzy rules obtained can create expert systems so that oncologists have more computerized tools to assist in the assessment and treatment of their patients. In addition to the introduction, this paper introduces the section of theoretical references (Section II), linked to the concepts of intelligent models and related work on breast cancer. In section III, the model used to solve the problem raised in the paper will be presented to the reader, as well as the experiments, and their results will be highlighted to the reader in section IV. Finally, future work and conclusions about the activities performed in the paper will be presented in Section V.

II. Breast Cancer Concepts

A. Breast Cancer Treatments And Its Efficiency

Studies on the effectiveness and impacts of breast cancer survivors are diverse, exploring the emotional, body, and social aspects. They have been going on since the late 1980s and go to the present day. That has become an essential element in research due to the worldwide mortality rate among women and the exponential increase in cases in recent years [14].

A variety of conditions causes breast cancer — hereditary factors, advanced age, reproductive context, hormonal, and genetic. Of the highlighted factors, age is currently the factor that most correlates with cancer incidence [14]. However, it should be noted that there are behavioral factors that make it possible to reduce the risk of breast cancer. Breastfeeding and physical activity protect from pre and post-menopausal breast cancer [15]. Breast cancer can also originate due to uncontrolled exposure to ionizing factors.

With the evolution of science, treatments have become less mutilating to the female body, and the use of individual aspects to meet the specific needs of the women's body was designed to reduce pain and suffering during its execution. Several factors interfere with the treatment of breast cancer, mainly related to the time of its discovery. The breast cancer prognosis depends on the disease extent (stage) as well as the tumor characteristics (such as metastasis). The sooner the disease is discovered in women, the higher their chances of success in their treatment. Currently, the main treatments involve chemotherapy, radiotherapy, breast removal, and reconstruction surgery, hormone therapy, among others, [16]. The type of treatment to use depends on the stage of breast cancer, and the follow are the main stages and their correlated treatment approaches [16]:

Stages I and II: Surgery usually occurs with the removal of the tumor or breast. The applied method can also address the maintenance of breast and nipple epithelial tissue to facilitate breast reconstruction.

Stage III: They have localized tumors that are simpler to identify because of their increased size. The approach under these circumstances is linked to initial chemotherapy treatment.

Stage IV: Treatment should balance the tumor response to the most aggressive treatments and the possible prolongation of patient survival, taking into account the potential side effects of treatment.

B. Related works

The effectiveness of cancer treatment and patient survival are factors commonly addressed by the scientific community during related research on breast cancer. The identification of correlations, determining factors for the survival of a woman undergoing treatment, has become a preponderant element in leading researchers worldwide [17]. Research has sought to evaluate care for breast cancer survivors by identifying sociodemographic and medical aspects [18]. Other studies, as in the paper by Mcneely et al. [19] which addresses a systemic review of the influence of exercise on cancer treatment survivors. Studies in the 1990s already assessed the long-term impacts on patients who survived treatment by trials of radiotherapy [20]. Other studies have sought to understand the relationship between cognitive performance [4], lymphedema Symptoms measurement changes, and limb treatment [21] for survivors.

III. Autonomous Fuzzy Neural Network

This paper proposes an autonomous fuzzy neural network, as shown in Fig. 1. The FNN includes three layers where the first is responsible for the fuzzification process, the second build fuzzy rules, and finally, the third represents a neural aggregation network. The main modification proposed in the model is the use of an autonomous approach to the definition of fuzzy neurons in the first layer of the model, acting as an independent approach to fuzzification.

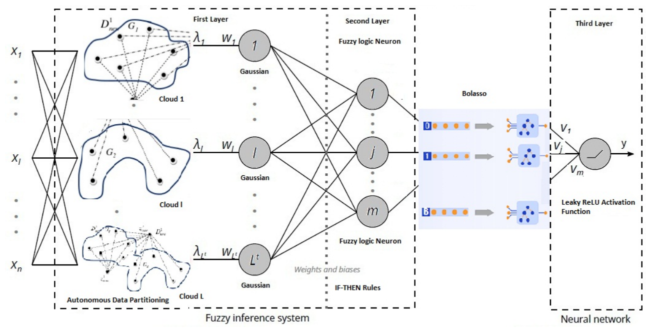

Fig. 1. Autonomous Fuzzy Neural Network Architecture

That is the first autonomous fuzzification hybrid model to act as a fuzzy inference system in breast cancer problems. Thus the rules found will be entirely based on the nature of the problem data.

The following subsections will present the concepts involved in the architecture of the constructed model represented in Fig. 1.

A. First Layer: Autonomous Fuzzification Method

The first layer of the proposed model is composed of fuzzy neurons formed by autonomous fuzzifying the problem's samples. The use of fuzzy sets is used to determine the degree of feature relevance of the problem to get the expected outputs. Fuzzy processes applied to the input layer of fuzzy models make it possible to create inference systems based on logical rules. The most traditional approaches are Grid-type partitioning, Clustering-based partitioning, and GA-based partitioning [22]. This step provides fuzzy inputs to one for the second layer of the model by processing a set of IF-THEN rules, which are the knowledge representation through linguistic expressions. Therefore, choosing a proper fuzzification technique can allow only meaningful rules to be generated about the problem evaluated. If the fuzzification process is chosen incorrectly, model processing can lead to sizeable computational processing.

The fuzzy neural network proposed in this paper considers the fuzzification process through a non-parametric technique capable of autonomously representing the maximum data density locations, thus allowing the simple construction of clusters representing fuzzy neurons in the first layer. Therefore the process of handling model inputs can be viewed as an autonomous, non-parametric procedure that is centered on the nature of the data. That ensures that the fuzzy rules generated correspond to the data to be evaluated, generating assertive answers with a high degree of interpretation of the problem. For each input variable (x_{ij}), a number of clouds are defined A_{lj}, l = 1... L, whose any clouds are the activation functions of the corresponding neurons. Thus, the outputs of the first layer are the membership degrees associated with the input values, i.e., a_{jl} = μ_l^A for j = 1... N and l = 1 ...L, where N is the number of inputs and L is the number of clouds for each input.

1) Autonomous fuzzification process: The ADPA technique [23] was inspired by the approach proposed by Angelov et al. [24], called the empirical data analysis (EDA) framework. Therefore, the fuzzification method used in this paper addresses an evolution in traditional clustering techniques, making them corresponding to the complexity of the data analyzed and less parametric, since there is no need to define the number of divisions or clusters needed to fuzzify the inputs pace. For this, the technique acts on the problem data by evaluating four recursively calculated parameters: the cumulative proximity, the standardized eccentricity, the density, and the multimodal typicality of the data evaluated in the problem [25].

EDA considers the input variables of the intelligent model as **x** = $\{x_1, x_2, ..., x_k\} \in \mathbb{R}^d$, where the indices k indicate the time instance at which the data point arrives. It can define the set of unique data point locations in time instance k as **u**= $\{u_1,\ u_2,\ ...,\ u_l,\} \in \mathbb{R}^d$ and the corresponding number of times that $f_1, f_2, ..., f_l$ different data points occupy the same unique locations. Based on unique data point locations, it is possible to reconstruct the data set $x_1, x_2, ..., x_k$ exactly if necessary, regardless of the order of arrival of the data points [25]. Based on this concept, the ADPA technique uses distance, proximity, and formation of data clouds to involve the concepts of cumulative proximity ((π_k^u)) for the formation of data clusters and, consequently, to determine the fuzzy procedure of the fuzzy neural network. Mathematically, we can write it as (1), where the distance (this paper considered the Chebyshev distance [26]) between two single sites is defined by $d(u_i, u_j)$ [25]:

$$\pi_k^u(u_i) = \sum_{j=1}^{l} max\ |u_i - u_j|, i = 1, 2, \dots, l \tag{1}$$

where l is the number of the unique data samples. Using the definitions of Eq. 1, it defines local density as Eq. (2) [25].

$$D_k(x_k) = \frac{\sum_{j=1}^{k}\sum_{l=1}^{k} max\ |u_i - u_j|}{2k\sum_{l=1}^{k} max\ |u_i - u_j|} \tag{2}$$

The standardized eccentricity (ε_k^u), and for the data point locations (**u**) can be calculated by Eq. (3) [25].

$$\begin{gathered}\varepsilon_k^u(u_i) = \frac{2\pi_k^u(u_i)}{E[\pi_k^u(u)]}, E[\pi_k^u(u)] > 0;\\ k > 1; l > 1\ and\ i = 1, 2, \dots, l;\end{gathered} \tag{3}$$

where $E[\pi_k^u(u)] = \frac{1}{l}\sum_{j=1}^{l}\pi_k^u(u_j)$ represents the mean cumulative proximity [27].

Data density (D_k^u) (inverse of standardized eccentricity), and typicality (when data density is normalized) (τ_k^u) are defined in (4) and (5), respectively, and calculated as described in [25].

$$D_k^u(u_i) = \frac{1}{\varepsilon_k^u(u_i)} \tag{4}$$

$$\tau_k^u(u_i) = \frac{D_k^u(u_i)}{\sum\limits_{j=1}^{1} D_k^u/(u_j)}, \sum_{j=1}^{l} D_k^u(u_j) > 0, k > 1, l > 1 \tag{5}$$

Finally, an essential concept for the EDA calculation is the multimodality typicality (τ_k^{MM}) and is mathematically defined as Eq.(6) [25].

$$\tau_k^{MM}(u_i) = \frac{f_i \tau_k^u(u_i)}{\sum\limits_{j=1}^{l} f_j \tau_k^u(u_j)} = \frac{f_i D_k^u(u_i)}{\sum\limits_{j=1}^{l} f_j D_k^u(u_j)}$$
$$\text{where } \sum_{j=1}^{l} f_j \tau_k^u(u_j) > 0, \sum_{j=1}^{l} f_j D_k^u(u_j) > 0, k > 1, l > 1 \tag{6}$$

EDA measures are responsible for operating the functions performed in the ADPA algorithm, directly acting on the properties of the observed dataset, releasing it from the need to use previous data on the data generation model and problem-specific parameters, thus generating a technique of less complicated and more interpretable FNN [23]. For the use of the ADPA algorithm, it is necessary to identify global density (D_n^G), and this factor is defined for unique data samples together with their corresponding numbers of repetitions in the dataset/stream, and of a particular unique data sample, u_i ($i = 1, 2, ..., n_u$; $n_u \geq 1$) is expressed as the product of its data density and its number of repetitions considered as a weighting factor as expressed in Eq. (7).

$$D_n^G(u_i) = f_i D_n(u_i) \tag{7}$$

The ADPA method can use three approaches that can to construct a cloud data pool. In this paper, we use the evolving approach, initiating its steps with a single sample, when the algorithm learns from a sample-by-sample basis, thus allowing the model to start its procedure without prior knowledge of the data. The approach has three main steps [23]: when initializing the data cloud, and it is responsible for selecting the first sample in a large data stream. ADPA performs the evolution of parameters and distances from the insertion of new samples in this context [23]. In sequence, ADPA recursively updates the new data evaluated by the model and defines the system structure. and finally, the multimodal typicality (τ_{k+1}^{MM}) can be updated using the expression shown in (8), as well as the cumulative proximity (π_{k+1}) through Eq. (9) [24].

$$\tau_{k+1}^{MM}(u_i) = \begin{cases} \frac{f_i D_k^u(u_i)}{\sum\limits_{j=1}^{l} f_j D_k^u(u_j) + D_k^u(u_i)} & u_i \neq x_{k+1}, u_j = x_{k+1} \\ \frac{f_i D_k^u(u_i)}{\sum\limits_{j=1}^{l} f_j D_k^u(u_j) + D_k^u(u_i)} & u_i = x_{k+1} \end{cases} \tag{8}$$

$$\pi_{k+1}^u(u_i) = \pi_k^u(u_i) + d^2(u_i, u_{+1}), i = 1, 2, \dots l \tag{9}$$

Since the ADPA technique uses the Chebyshev distance, one must consider the cumulative proximity updating as Eq. (10) and Eq. (11) [24].

$$\begin{aligned} \pi_{k+1}^u(u_{l+1}) = (l+1)\left((\mathbf{u}_{+1} - \varrho_{l+1})^{\mathrm{T}}\right. \\ \left.(u_{l+1} - \varrho_{l+1}) + u_{l+1} - \varrho_{l+1}^T \varrho_{l+1}\right) \end{aligned} \tag{10}$$

where $\varrho_{l+1} = \frac{l}{l+1}\varrho_l + \frac{1}{l+1}\mathbf{u}_{+1}$, and:

$$\mathbf{u}_{+1} = \frac{l}{l+1}\mathbf{u} + \frac{1}{l+1} u_{l+1}^{\mathrm{T}} u_{l+1} \tag{11}$$

Using Eq. (9), Eq. (10) and Eq. (11), it is possible to define the local density recursively too, like in Eq. (12) [23].

$$D_k(x_k) = \frac{1}{1 + \frac{\|x_k - \varrho_k^2\|}{x_k - \|\varrho_k^2\|}} \tag{12}$$

The local density and the centers of all the existing data clouds are calculated using recursive local density Eq. (10). In this paper, as in Gu et al. [23], it uses C_k as the number of existing local modes at the k time instance. When using the recursive concepts to the definition of the clouds, it is necessary to evaluate the results through the rule Eq. (13) [23], [28] to will form a new data cloud.

$$\begin{aligned} IF\ \left(D_k(x_k) > \overset{C_k}{\underset{i=1}{max}}\left(D_k\left(\varrho_k^i\right)\right)\right) \\ OR\ \left(D_k(x_k) < \overset{C_k}{\underset{i=1}{min}}\left(D_k\left(\varrho_k^i\right)\right)\right) \\ THEN\ (x_k\ becomes\ a\ new\ focal\ point) \end{aligned} \tag{13}$$

If this condition is satisfied, a new data cloud will be added with x_k as its local mode. Next, the method is responsible for the formation of data clouds, where local peaks (Θ) identified in the index list are used to attract the data samples that are closest to them using a $\min$ operator defined as Eq. (14) [23].

$$C_{win} = \underset{j=1,...,l}{\arg\min}(\|x_i - \Theta_j\|);\ i = 1, ..., k; l > 1 \tag{14}$$

Several Voronoi tessellations [29] are naturally formed, assigning all problem data samples to the nearest local maxima, and data clouds are built around local maxima. If the condition is satisfactory, a new data cloud is added with x_k in its local mode. That means that the corresponding cloud C_k is updated as in Eq. (15) [23].

$$\begin{aligned} C_k \leftarrow C_k + 1 \\ \varrho_k^{C_k} \leftarrow x_k \\ S_k^{C_k} \leftarrow 1 \end{aligned} \tag{15}$$

After forming the data clouds, the actual centers (mean), standard deviation per data cloud, and the supports (S_k), can be easily calculated. Besides, a rule-based check defines the progress of the assignment of the evaluated data to its representation space using the condition in Eq. (16).

$$IF\left(x_k - \varrho_k^n \leq \frac{\Delta_k^c}{2}\right)\ THEN\ (x_k\ is\ assigned\ to\ \varrho_k^n) \tag{16}$$

where ϱ_k^n represents local mode closest and Δ_k^c is the average Chebyshev distance between any pair of existing local modes [23]. If the condition established in Eq. (16) is fulfilled, then the evaluated sample is associated with the nearest existing local mode, and the model meta-parameters are updated as depicted in Eq. (17) and Eq. (18). Otherwise, the updating of the model follows Eq. (15).

$$S_k^n \leftarrow S_k^n + 1 \tag{17}$$

$$\varrho_k^n \leftarrow \frac{S_k^n - 1}{S_k^n}\varrho_k^n + \frac{1}{S_k^n}x_{k+1} \tag{18}$$

Finally, in the third and last step, if there are no more data samples to be grouped, the identified local modes are used to build data clouds using Eq. (14). Therefore the cloud creation procedure works with each new sample submitted to the model, thus allowing problems that have volatility in their main characteristics to be thoroughly evaluated by the clustering algorithm. Thus the evolving approach allows the creation of representative neurons in the first layer, being such responsible for the fuzzification process of the FNN, allowing the hybrid model proposed in this paper to have high adaptability to problems of complex nature. Figure 2 shows an example of the identified centers created in the fuzzification process in the databases evaluated in this paper (age-race relationship).

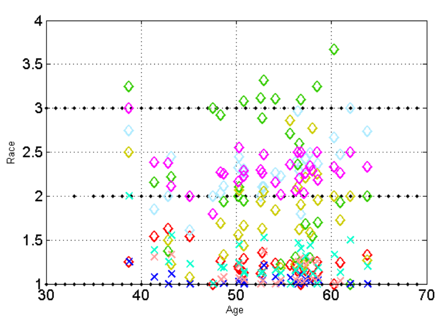

Fig. 2. ADPA method performed between the age-by-race relationship of women with breast cancer.

B. Second Layer: Fuzzy Rules

The second layer of the model consists of neurons composed of fuzzy equations, where the inputs of these equations are the fuzzy sets formed in the first layer of the model, aggregated with randomly defined synaptic weight values in the range between zero and one. Fuzzy rules are primarily responsible for receiving the knowledge of a database and turning it into logical relationships through IF-THEN rules. In this paper, the fuzzy rule should be seen as a Unineuron type III fuzzy neuron [30], which uses the fuzzy set operator as the uninorm [31]. The uninorm used in this paper is as follows:

$$U(x,y) = \begin{cases} o\,T(\frac{x}{o}, \frac{y}{o}), if\ y \in [0, o] \\ o + (1-o)S(\frac{x-o}{1-o}, \frac{y-o}{1-o}), if y \in [o, 1] \\ \delta(x,y), otherside.\ Where \\ \delta(x,y) = \begin{cases} max(x,y) & -\text{ if } \ g \in [0,\ 0.5] \\ min(x,y) & -\text{ if } \ g \in (0.5,\ 1] \end{cases} \end{cases} \tag{19}$$

where T is a *t-norm* (probabilistic sum), S is a *s-norm* (product) and o is the activation of the fuzzy neuron (defined randomly in the interval between zero and one).

The unineuron proposed in Lemos et al. [30] performs operations to compute its output through each pair $(a_i,\ w_i)$ is transformed into a single value b_i = **h** $(a_i,\ w_i)$ and calculate the unified aggregation of the transformed values **U** $(b_1, b_2 ... b_n)$ through the function p (relevancy transformation), that is responsible for transforming the inputs and corresponding weights into individual transformed values and $\bar{w}$ represents the complement of w. A formulation for the p function can be described as [30]:

$$p(w;a;o) = wa + \bar{w}o \tag{20}$$

using the weighted aggregation reported above the unineuron can be written as [30]:

$$\mathbf{z} = UNI(w;a;o) = U_{i=1}^{n} p(w_i, a_i, o) \tag{21}$$

For each **z**, a fuzzy equation is generated capable of extracting knowledge from the database in relational form:

$$\begin{aligned} Rule_z:\ IF\ x_{i1}\ is\ A_1^1\ with\ certainty\ w_{11}... \\ THEN\ z_z\ =\ v_z \end{aligned} \tag{22}$$

C. Third Layer: Neural Aggregation Network

The third layer of the model represents an aggregation neural network composed of a neuron responsible for the outputs of the model. The output of the model is:

$$y = \beta \sum_{j=0}^{l} f_{LeakyReLU}(z_l, v_l) \tag{23}$$

where z_0 = 1, v_0 is the bias, and z_j and v_j, j = 1, ..., l are the output of each fuzzy neuron of the second layer and their corresponding weight, respectively. Leaky ReLUs are one attempt to fix the ReLU problem. Instead of the function being zero when x $<$ 0, a leaky ReLU will instead have a small negative slope (considered in this paper with the value 0.01). That is, the function computes $f(x) = \Bbbk(x<0)(\alpha x) + \Bbbk(x>=0)(x)$ where α is a small constant. $f_{LeakyReLU}$ therefore, it is defined as: [32]:

$$f_{LeakyReLU}(\mathbf{x}, \alpha) = max(\alpha\mathbf{x}, \mathbf{x}) \tag{24}$$

The β function acts similarly to the sign function, which returns value 1 if the analyzed number is positive and -1 if it is negative. Thus the β function is described as:

$$\beta = \begin{cases} 1, \ if \ f_{LeakyReLU}(z_l, v_l) \ > 0 \\ -1, \ if \ f_{LeakyReLU}(z_l, v_l) \ < 0 \end{cases} \quad (25)$$

This argument is critical to defining binary pattern classification problems, such as what is being analyzed in this paper.

D. Model Training

For the determination of the weights that bind the second and third layers, we use the concepts of pseudo inverse [33] for the determination of the weights and is based on the training techniques proposed by Huang et al. [34]:

$$\mathbf{v} = \mathbf{Z}^{+}\mathbf{y} \quad (26)$$

where $\mathbf{Z}^{+}$ is pseudo-inverse of Moore-Penrose [35], of $\mathbf{z}$. The fuzzification technique used in the model can create an undesirable quantity of neurons and interfere with training overfitting, generating incorrect outputs. To avoid this problem, this model uses a re-sampling regularization technique to determine the most significant neurons. Using replications and crossing the results obtained during a trait selection criterion, a consistent estimation of the model is obtained, without the condition that by regression methods commonly used for this purpose, for example, by the regular Lasso. Therefore, Bolasso's structure [36] can be viewed as a voting scheme applied to the supports of the initialization Lasso estimates. [2] However, the procedure should be viewed as a consensus matching scheme, as the maximum relevant subset of the problem variables with which all regressors agree on predefined selection terms remains, making the method proven to be a more consistent statistical approach.

When it evaluate a set of N distinct samples (x_i, y_i), where $x_i = [x_{i1}, x_{i2} \dots x_{id}]^T \in \mathbb{R}^d$ where d is the dimension of x_i and $y_i \in \mathbb{R}$ for $i = 1...N$, the cost function of this regression algorithm can be defined as:

$$\sum_{i=1}^{N} \parallel \mathbf{z}(x_i)v - y_i \parallel_2 + \lambda \parallel v \parallel_1 \quad (27)$$

Where λ is a regularization parameter usually estimated using cross-validation [37]. The LARS algorithm can be used to select the model already. [3] For the problem examined in this paper, the z_l regressors are the results of meaningful neurons.

IV. Experiment

A. Dataset used in the test

The study database was based on the work of Teng et al. [6], which worked based on summary statistics of breast cancer patients as of the November 2018 update of the NCI SEER Program. All samples collected were from females (5,279 patients) with infiltrating duct and breast cancer with lobular carcinoma (NOS histology codes 8522/3) diagnosed with the disease within five years (2006-2010). Assumptions of choosing which patients to include in the data analysis were defined by Teng et al. [6]; thus, they were finally included in our analysis. The main problem variables are listed below: [4] *Age* = numerical; *Race* (White, Black and Other.); *Marital status* (Divorced, Married, Separated, Single and Widowed.); *T stage* (T1, T2, T3, T4.); *N stage* (N1, N2 and N3.); *6th stage* (IIA, IIB, IIIA, IIIB, IIIC.); *Grade* (Grade I, Grade II, Grade III, Grade IV.); *A stage* (Regional and Distant.); *Tumor size* (<36 mm, 36 – 70 mm, 71 – 105 mm, >105 mm.); *ER status and PR Status*(Positive and Negative.); *Regional nodes* (Total and Positive.); *Survival months*=numeric; *Status* (Alive and Dead.).

To meet the assumptions of the model proposed in this paper, all non-numeric values were converted to sequential numbers starting from 1. Thus, Marital status is now represented by 1, 2, 3, 4, and 5, respectively. The same happens when the attributes are Positive (1) and Negative (2) and so on with all dimensions of the problem.

B. Models and test premisses

The fuzzy neural network used in the experiments (FNN) had its hyperparameters defined by preliminary procedures[5], where the parameters were defined through the best result of the pattern classification test, accompanied by the highest AUC. To make a pertinent comparison on the ability to identify people who can survive cancer treatment, smart techniques will be used to determine if the results of the hybrid approach meet the state of the art in pattern classification. Those models are available in the tool WEKA [38] and them are Naive Bayes (NB) [6], Bayes Net (BN) [7], Random Forest (RF) [8], Multilayer Perceptron (MLP) [9] and C4.5 [10].

The factors evaluated in this paper are as follows:

$$acc = \frac{TP + TN}{TP + FN + TN + FP} \quad (28)$$

$$AUC = \frac{1}{2}(sens + spec) \quad (29)$$

[2] A consensus threshold is defined, say η = 70 %, and a regressor is included if selected in at least 70 % of assays.

[3] For a given value of λ, only a fraction (or none) of the regressors have corresponding nonzero weights. If λ = 0, the problem becomes unrestricted regression, and all weights are nonzero. As λ increases from 0 to a given value λ_{max}, the number of nonzero weights decreases until it reaches 0.

[4] All non-numeric attributes were converted to sequential numbers so that the algorithms could act equally.

[5] using the 10 k-fold technique for the values of η = [0.5, 0.6, 0.7], the number of bootstrap replications = [8, 16, 32] and σ = [10, 20, 30, 40, 50] for Gaussian neurons in the first layer of the model. The results were obtained by training/testing partitions (70/30)

[6] useKernelEstimator=false, debug=false, displayModelInOldFormat=false, doNotCheckCapabilities=false, useSupervisedDiscretization=false

[7] estimator=SimpleEstimator, debug=false, searchAlgorithm =F2, doNotCheckCapabilities=false, useADTree=false

[8] seed=1, allowUnclassifiedInstances=false, debug=false, minNum=1.0, numFolds=0, doNotCheckCapabilities=false, maxDepth=0, minVarianceProp=0.001, KValue=0

[9] In all tests the following settings were used: batch size = 100, hidden layers = 1, learning rate = 0.3, momentum = 0.2, validation Threshold = 20.

[10] seed=1, unpruned=false, confidenceFactor=0.25, numFolds =3, reducedErrorPruning=false, useLaplace=false, doNotMakeSplitPointActualValue=false, debug=false, subtreeRaising=true, saveInstanceData=false, binarySplits=false, doNotCheckCapabilities=false, minNumObj=2, useMDLcorrection=true

In this context, consider specificity and sensitivity as:

$$sens = \frac{TP}{TP + FN} \tag{30}$$

$$spec = \frac{TN}{TN + FP} \tag{31}$$

where, $TP =$ true positive, $TN =$ true negative, $FN =$ false negative and $FP =$ false positive.

C. Test results

Test results are presented (Table I) after performing 30 repetitions with randomly arranged samples for all models involved in the test. It should be noted that the values presented refer to the averages of the 30 measurements and that the values in parentheses represent the standard deviation, a relevant factor to evaluate the stability of the model during the tests performed. Finally, the time of the experiment is presented in seconds. The bolded values of bold represent the best results obtained in the test. All tests were performed on a computer with the following settings: Core (TM) 2 Duo CPU, 2.27 GHz with 3-GB RAM.

TABLE I
SURVIVAL ASSESSMENT RESULTS FOR BREAST CANCER TREATMENTS.

Model	Acc.	AUC	Spec.	Sens.	Time (sec)
FNN	**90.46 (1.02)****	**0.7428 (0.02)**	96.41 (0.02)	**52.16 (2.98)**	55.71 (9.82)
BN	84.83 (1.04)	0.7219 (0.03)	94.41 (0.01)	49.98 (0.03)	**0.01 (0.00)**
MLP	87.19 (0.78)*	0.7055 (0.01)	94.52 (0.01)	46.59 (0.03)	61.83 (1.63)
NB	83.78 (0.25)	0.7059 (0.03)	89.60 (0.01)	51.58 (0.03)	0.02 (0.00)
C4.5	90.30 (0.76)*	0.7308 (0.01)	**97.90 (0.01)**	48.26 (0.04)	0.69 (0.01)
RT	83.84 (0.88)	0.6865 (0.01)	90.53 (0.01)	46.78 (0.04)	0.29 (0.00)

The model proposed in this paper obtained the best numerical results in the classification accuracy of survivors of cancer treatment, as well as having better indices in the system's ability to correctly predict the situation for cases that have a chance of surviving cancer treatment and its outcome. [11] AUC index was also the highest of the experiments performed. According to Teng et al. [6], AUC results above 0.7 indicate a good scoring model for identifying mortality in people who underwent breast cancer treatment. Note that the low sensitivity in all tests is due to a very unbalanced dataset. TTherefore we can say that the approach brings interesting results to science and, at the same time, can extract fuzzy rules with high interpretability.

D. Fuzzy rules generated

The main advantage of using fuzzy neural network models is their ability to extract knowledge from a database. Therefore, the following rules were extracted from the experiment, and the ability provides their relevance to results adapt to a linguistic way. Two fuzzy rules are presented below in their numerical version and through linguistic relations. That allows you to check more clearly the relationships within the problem.

If *Age* = 0.5936 and/or *Race* = 0.0409 and/or *Marital status* = 0.5308. and/or *T stage* = 0.6904 and/or *N stage*= 0.6829 and/or *6th stage*= 0.7428 and/or *Grade*=-0.06303 and/or *A stage* = 0.3628 and/or *Tumor size*=0.7829 and/or *ER status*= 0.4100 and/or *PR status* = 0.5069 and/or *Regional nodes*= 0.4708 and/or *Survival months*= 0.2738 Then *Status*= 0.6872.

If *Age* = -0.2608 and/or *Race* = 0.0142 and/or *Marital status* = 0.7096 and/or *T stage* = 1.1272 and/or *N stage*= 0.7755 and/or *6th stage*= 1.1264 and/or *Grade*= -0.5388 and/or *A stage* = 0.3712 and/or *Tumor size*=0.9158 and/or *ER status*= 0.4946 and/or *PR status* = 0.7893 and/or *Regional nodes*= 0.5812 and/or *Survival months*=0.8525 Then *Status*= -0.0623

If *Age* = very old and/or *Race* = white and/or *Marital status* = Married and/or *T stage* = T1 and/or *N stage*= N2 and/or *6th stage*= IIA and/or *Grade*=II and/or *A stage* = Regional and/or *Tumor size*=Medium and/or *ER status*= Positive and/or *PR status* = Positive and/or *Regional nodes*= Positive and/or *Survival months*= shortly Then *Status*= Alive

If *Age* = very young and/or *Race* = white and/or *Marital status* = Separeted and/or *T stage* = T3 and/or *N stage*= N2 and/or *6th stage*= IIIA and/or *Grade*= I and/or *A stage* = Regional and/or *Tumor size*=Large and/or *ER status*= Positive and/or *PR status* = Positive and/or *Regional nodes*= Positive and/or *Survival months*=long time Then *Status*= Dead.

V. CONCLUSION

The results obtained in the pattern classification tests regarding mortality in women undergoing breast cancer treatment confirm that the fuzzy neural network addressed in this paper can act efficiently in determining the possibility of cancer treatment survival. The results are equivalent to models traditionally used in the literature and are also better than the original study proposed by Teng et al. [6]. Besides, fuzzy rules have been extracted from the problem and can be seen as an interpretable relationship between the problem dimensions. These fuzzy rules can be seen as the basis for building an inference system that can assist the routine of medical oncologists and for women who want to assess dangerous situations after undergoing cancer treatment.

Future work may address other intelligent models, new hybrid architectures, statistical approaches to define better model parameters, as well as the construction of an expert system through the fuzzy rules obtained in this work.

ACKNOWLEDGMENT

The authors acknowledge the support by the Austrian Science Fund (FWF): contract number P32272-N38, acronym IL-EFS.

[11] All test results were evaluated using a statistical test (ANOVA) [39]. With a 95% probability, we can say that all results marked with an '*' are statistically equal concerning the equitable performance of the factors collected between the models analyzed in the test. All premises (normality of residues, homoscedasticity, and independence) were not violated.

References

[1] J. R. Benson and I. Jatoi, "The global breast cancer burden," *Future oncology*, vol. 8, no. 6, pp. 697–702, 2012.

[2] L. J. Fallowfield, S. K. Leaity, A. Howell, S. Benson, and D. Cella, "Assessment of quality of life in women undergoing hormonal therapy for breast cancer: validation of an endocrine symptom subscale for the fact-b," *Breast cancer research and treatment*, vol. 55, no. 2, pp. 187–197, 1999.

[3] A. R. Costa, F. Fontes, S. Pereira, M. Gonçalves, A. Azevedo, and N. Lunet, "Impact of breast cancer treatments on sleep disturbances–a systematic review," *The Breast*, vol. 23, no. 6, pp. 697–709, 2014.

[4] J. E. Carroll, K. Van Dyk, J. E. Bower, Z. Scuric, L. Petersen, R. Schiestl, M. R. Irwin, and P. A. Ganz, "Cognitive performance in survivors of breast cancer and markers of biological aging," *Cancer*, vol. 125, no. 2, pp. 298–306, 2019.

[5] P. Minicozzi, L. Van Eycken, F. Molinie, K. Innos, M. Guevara, R. Marcos-Gragera, C. Castro, E. Rapiti, A. Katalinic, A. Torrella *et al.*, "Comorbidities, age and period of diagnosis influence treatment and outcomes in early breast cancer," *International journal of cancer*, vol. 144, no. 9, pp. 2118–2127, 2019.

[6] J. Teng, A. Abdygametova, J. Du, B. Ma, R. Zhou, Y. Shyr, and F. Ye, "Bayesian inference of lymph node ratio estimation and survival prognosis for breast cancer patients," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2019.

[7] P. V. C. Souza, A. G. dos Reis, G. R. R. Marques, A. J. Guimaraes, V. J. S. Araujo, V. S. Araujo, T. S. Rezende, L. O. Batista, and G. A. da Silva, "Using hybrid systems in the construction of expert systems in the identification of cognitive and motor problems in children and young people," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, June 2019, pp. 1–6.

[8] C.-W. Han, "Fuzzy neural network-based fetal health monitoring using cardiotocography data," in *AIP Conference Proceedings*, vol. 2104, no. 1. AIP Publishing, 2019, p. 030008.

[9] Y. Liu, S. Wu, Kuang-Pen Chou, Y. Lin, Jie Lu, Guangquan Zhang, Wen-Chieh Lin, and C. Lin, "Driving fatigue prediction with pre-event electroencephalography (eeg) via a recurrent fuzzy neural network," in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, July 2016, pp. 2488–2494.

[10] P. V. d. C. Souza and A. J. Guimaraes, "Using fuzzy neural networks for improving the prediction of children with autism through mobile devices." in *2018 IEEE Symposium on Computers and Communications (ISCC)*, June 2018, pp. 01 086–01 089.

[11] A. J. Guimarães, V. J. S. Araujo, V. S. Araujo, L. O. Batista, and P. V. de Campos Souza, "A hybrid model based on fuzzy rules to act on the diagnosed of autism in adults," in *Artificial Intelligence Applications and Innovations*, J. MacIntyre, I. Maglogiannis, L. Iliadis, and E. Pimenidis, Eds. Cham: Springer International Publishing, 2019, pp. 401–412.

[12] P. V. de Campos Souza, A. J. Guimaraes, V. S. Araujo, T. S. Rezende, and V. J. S. Araujo, *Using Fuzzy Neural Networks Regularized to Support Software for Predicting Autism in Adolescents on Mobile Devices*. Singapore: Springer Singapore, 2019, pp. 115–133. [Online]. Available: $https://doi.org/10.1007/978-981-13-8614-5_7$

[13] V. J. Silva Araújo, A. J. Guimarães, P. V. de Campos Souza, T. Silva Rezende, and V. Souza Araújo, "Using resistin, glucose, age and bmi and pruning fuzzy neural network for the construction of expert systems in the prediction of breast cancer," *Machine Learning and Knowledge Extraction*, vol. 1, no. 1, pp. 466–482, 2019.

[14] C. E. DeSantis, J. Ma, M. M. Gaudet, L. A. Newman, K. D. Miller, A. Goding Sauer, A. Jemal, and R. L. Siegel, "Breast cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, 2019.

[15] A. C. Peterson and H. Uppal, "Method for predicting response to breast cancer therapeutic agents and method of treatment of breast cancer," Feb. 5 2019, uS Patent 10,196,693.

[16] C. I. Li, K. E. Malone, and J. R. Daling, "Differences in breast cancer stage, treatment, and survival by race and ethnicity," *Archives of internal medicine*, vol. 163, no. 1, pp. 49–56, 2003.

[17] M. Cappiello, R. S. Cunningham, M. Tish Knobf, and D. Erdos, "Breast cancer survivors: information and support after treatment," *Clinical nursing research*, vol. 16, no. 4, pp. 278–293, 2007.

[18] E. E. Pauwels, C. Charlier, I. De Bourdeaudhuij, L. Lechner, and E. Van Hoof, "Care needs after primary breast cancer treatment. survivors' associated sociodemographic and medical characteristics," *Psycho-oncology*, vol. 22, no. 1, pp. 125–132, 2013.

[19] M. L. McNeely, K. L. Campbell, B. H. Rowe, T. P. Klassen, J. R. Mackey, and K. S. Courneya, "Effects of exercise on breast cancer patients and survivors: a systematic review and meta-analysis," *Cmaj*, vol. 175, no. 1, pp. 34–41, 2006.

[20] J. Cuzick, H. Stewart, L. Rutqvist, J. Houghton, R. Edwards, C. Redmond, R. Peto, M. Baum, B. Fisher, and H. Host, "Cause-specific mortality in long-term survivors of breast cancer who participated in trials of radiotherapy." *Journal of Clinical Oncology*, vol. 12, no. 3, pp. 447–453, 1994.

[21] J. M. Armer, K. V. Ballman, L. McCall, N. C. Armer, Y. Sun, T. Udmuangpia, K. K. Hunt, E. A. Mittendorf, D. R. Byrd, T. B. Julian *et al.*, "Lymphedema symptoms and limb measurement changes in breast cancer survivors treated with neoadjuvant chemotherapy and axillary dissection: results of american college of surgeons oncology group (acosog) z1071 (alliance) substudy," *Supportive Care in Cancer*, vol. 27, no. 2, pp. 495–503, 2019.

[22] C.-F. Juang and C.-T. Lin, "An online self-constructing neural fuzzy inference network and its applications," *IEEE transactions on Fuzzy Systems*, vol. 6, no. 1, pp. 12–32, 1998.

[23] X. Gu, P. P. Angelov, and J. C. Príncipe, "A method for autonomous data partitioning," *Information Sciences*, vol. 460, pp. 65–82, 2018.

[24] P. Angelov, "Outside the box: an alternative data analytics framework," *Journal of Automation Mobile Robotics and Intelligent Systems*, vol. 8, no. 2, pp. 29–35, 2014.

[25] P. Angelov, X. Gu, and D. Kangin, "Empirical data analytics," *International Journal of Intelligent Systems*, vol. 32, no. 12, pp. 1261–1284, 2017.

[26] D. Gottlieb and S. A. Orszag, *Numerical analysis of spectral methods: theory and applications*. Siam, 1977, vol. 26.

[27] P. P. Angelov, X. Gu, and J. C. Príncipe, "A generalized methodology for data analysis," *IEEE transactions on cybernetics*, no. 99, pp. 1–13, 2017.

[28] P. Angelov, *Autonomous learning systems: from data streams to knowledge in real-time*. John Wiley & Sons, 2012.

[29] M. I. Shamos and D. Hoey, "Closest-point problems," in *16th Annual Symposium on Foundations of Computer Science (sfcs 1975)*. IEEE, 1975, pp. 151–162.

[30] A. Lemos, W. Caminhas, and F. Gomide, "New uninorm-based neuron model and fuzzy neural networks," in *2010 Annual Meeting of the North American Fuzzy Information Processing Society*. IEEE, 2010, pp. 1–6.

[31] R. R. Yager and A. Rybalov, "Uninorm aggregation operators," *Fuzzy sets and systems*, vol. 80, no. 1, pp. 111–120, 1996.

[32] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, 2013, p. 3.

[33] P. V. C. Souza, "Regularized fuzzy neural networks for pattern classification problems," *International Journal of Applied Engineering Research*, vol. 13, no. 5, pp. 2985–2991, 2018.

[34] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[35] A. Albert, *Regression and the Moore-Penrose pseudoinverse*. Elsevier, 1972.

[36] F. R. Bach, "Bolasso: model consistent lasso estimation through the bootstrap," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 33–40.

[37] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[38] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[39] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.