Available online at www.elixirpublishers.com (Elixir International Journal)

Network Engineering

Elixir Network Engg. 110 (2017) 48206-48212

ISSN: 2229-712X

Structural Elucidation of Eudesmane Sesquiterpenes using GRNN and Scatter Plots

Taye Temitope Alawode[1] and Kehinde Olukunmi Alawode[2]

[1]Department of Chemical Sciences, Federal University Otuoke, Bayelsa State, Nigeria.

[2]Department of Electrical and Electronic Engineering, Osun State University, Osun State, Nigeria.

ARTICLE INFO

Article history:

Received: 11 July 2017;

Received in revised form:

1 September 2017;

Accepted: 11 September 2017;

Keywords

GRNN,

Scatter plots,

Structural elucidation,

^{13}C,

Eudesmane sesquiterpenes.

ABSTRACT

This study seeks to achieve a complete elucidation of structures of unknown Eudesmane sesquiterpenes from their ^{13}C values. The ^{13}C values for each of the fifteen (15) positions of the skeletons of the Eudesmane compounds were predicted using Generalized Regression Neural Network (GRNN). From the predicted ^{13}C values, GRNN and Scatter Plot methods were used to predict the substituents attached to each position on the skeleton of the Eudesmane compounds. Recognition of the test compounds ranged between 40 and 100%. GRNN and Scatter plots demonstrated great potential for use in the structural elucidation of unknown compounds from ^{13}C values.

Introduction

A major challenge faced by natural products chemists in the process of drug development is elucidation of the structure of isolated compounds. Sesquiterpenes are formed from countless biogenetic pathways and therefore produce several types of carbon skeletons [1,2]. This makes elucidation of their structures very challenging. Eudesmane-type compounds are one of the most representative skeletons of sesquiterpenes. Emerenciano and co-workers have developed and applied the expert system, SISTEMAT (based on artificial intelligence programs) in the elucidation of structures of several classes of compounds including sesquiterpenes [3,4], lactone sesquiterpenes[5], diterpenes [6] and triterpenes [7]. Oliveira and co workers [4] demonstrated the use of SISTEMAT in obtaining useful rules of ^{13}C spectral analysis and its use as an auxiliary tool in the process of structure elucidation for eudesmanes. The same work presented a review on the ^{13}C NMR data of eudesmanes. Part of the ^{13}C NMR chemical shift data used in our previous studies and the current one are obtained from this publication. The structure of any natural product is conventionally divisible into three sub-units: the skeletal atoms, heteroatoms directly bonded to the skeletal atoms or unsaturations between them and secondary carbon atoms, usually bonded to a skeletal atom through an ester or other linkages [5]. The skeletal structure common to all Eudesmane compounds is shown in Fig.1.

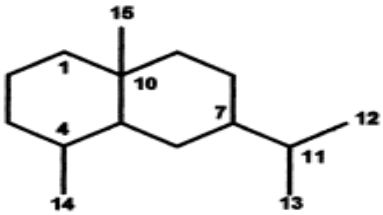

Fig 1. Eudesmane skeleton.

Typical substituents found in Eudesmane compounds are presented in Fig. 2.

Our Contribution

In a previous study [8], we have shown that when Generalized Regression Neural Network (GRNN), is trained using the ^{13}C chemical shift values for each of the 15 positions on the eudesmane skeleton as input and the various possible substituents as the target, GRNN could identify the substituents in each position on the Eudesmane skeleton of unknown compounds (when the ^{13}C values for each position on the Eudesmane skeleton of the each unknown eudesmane is supplied to the system). We have also applied scatter plot as a tool to determine the ^{13}C chemical shift ranges (for each of the 15 carbon positions on the Eudesmane skeleton) over which different substituent types may exist allowing the determination all the possible structures consistent with a particular set of spectroscopic data [9]. However, full elucidation of structures of unknown compounds using the described procedures could not be carried out since the studies were based on the premise that the ^{13}C skeletal data of the compounds whose substituents were being determined are known. In the present work, we predict the ^{13}C chemical shift values on the skeleton (C_1-C_{15}) of novel eudesmane skeleton using GRNN. We thereafter proceeded to utilize the predicted skeletal values to determine the substituent types in each position on the eudesmane skeleton employing the principles demonstrated in the previous works. The degree of accuracy of GRNN and scatter plots in determining the substituents on each position of the skeleton of the test compounds were compared. In utilizing the GRNN and scatter plots in predicting the substituent types, the original data set utilized in our previous publications were expanded in order to accommodate new substituent types encountered in the compounds employed in skeletal data prediction.

Tele: +234-7056712384

E-mail address: onatop2003@yahoo.com

Artificial Neural Networks

Artificial Neural Networks are computational models whose structures are derived from a simplified concept of the brain in which a number of nodes called neurons, are interconnected in a network-like structure [11]. Neural networks are non-linear processes that perform learning and classification. ANNs consist of a large number of interconnected processing elements known as neurons that act as microprocessors. Each neuron accepts a weighted set of inputs and responds with an output. Figure 2a depicts a single neuron model. In general, neural networks are adjusted/ trained to reach from a particular input a specific target output until the network output matches the target. Hence the neural network can learn the system. The learning ability of a neural network depends on its architecture and applied algorithmic method during the training. Training procedure ceases if the difference between the network output and desired/actual output is less than a certain tolerance value. Thereafter, the network is ready to produce outputs based on the new input parameters that are not used during the learning procedure.

A GRNN (an architecture of ANN) consists of four layers: input layer, pattern layer, summation layer and output layer as shown in Fig. 2b. The number of input units in the input layer depends on the total number of the observation parameters. The first layer is connected to the pattern layer and in this layer each neuron presents a training pattern and its output. The pattern layer is connected to the summation layer. The summation layer has two different types of summation, which are a single division unit and summation units. The theory of GRNN has been described elsewhere [12, 13].

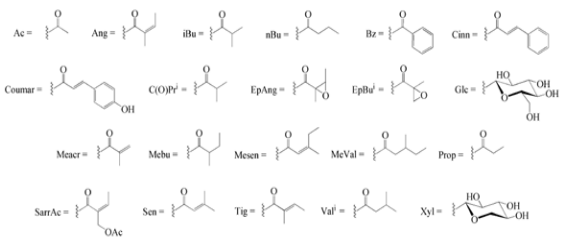

Fig 2. Common Substituents found in Eudesmane Compounds [10].

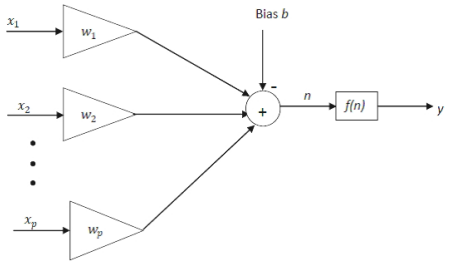

Fig 2a: A Single Neuron model [11].

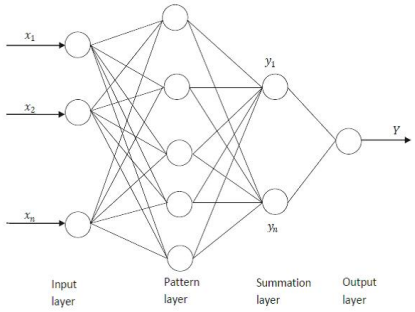

Fig 2b: General Structure of GRNN [12].

ANNs have been applied in the identification, distribution and recognition of patterns of chemical shifts from ^{1}HNMR spectra [14, 15] and identification of chemical classes through ^{13}C-NMR spectra [16].

Methodology

The ^{13}C spectral data for eudesmane sesquiterpenes used for the current study were extracted from structures of Eudesmane compounds published by [17]. Three Excel worksheets containing coded information on the input and target data for the training and test compounds were prepared. On the first row of the first sheet, the compounds were assigned codes 1-86. All the ^{13}C values for each compound were recorded in ascending order. In order to ensure equal number of entry for all the compounds, the difference in ^{13}C data was made up with zero for deficient compounds. This was used as the input data. The second sheet contains the target data and follows the format for the input data except that in the first column, the positions of each carbon atoms on the skeleton (as shown in Figure 1) were coded as 1-15 and the ^{13}C chemical shift data for each Carbon at each of the 15 positions was recorded for each compound. The third excel sheet contains the test compounds. The design follows that of the first sheet except that thirty-three (33) test compounds were utilized (coded 1-33). Owing to space limitations, only the ^{13}C chemical shift data of ten of the test compounds are given in Table 1. A list of these (ten) compounds (subsequently assigned codes 1-10) is provided in Appendix A.

The first step was to determine if there were discernable differences between the distribution of the ^{13}C data of whole compounds and those of their skeletons. This is to determine if there are specific ranges of ^{13}C chemical shift for eudesmane skeletons thereby allowing quick elimination of the ^{13}C data due to the substituents. To ascertain this, 3D-surface plots of the data on Excel Sheets 1 and 2 were carried out and compared (Figs 4a&b). It was observed from the plots that the ^{13}C data of the substituents are closely related to those of the eudesmane skeletons making the distinction between both sets of values difficult. This led to the exploration of artificial intelligence methods in the prediction of possible ^{13}C values for each of the 15 positions of the eudesmane skeleton. GRRN has been previously employed successfully by the authors to predict the substituents on the eudesmane skeleton [8]. It was therefore used in the current study. The data were transferred into the Neural Network toolbox of MATLAB 7.8.0 [18]. From the command window, the 'nntool' command was used to designate the imported data appropriately as 'input' or 'target' and to select the appropriate network for training. GeneralizedRegression Neural networks (GRNN) was selected for the training and subsequent simulation of the test compounds. Least deviation of ^{13}C values from those of the test compounds was obtained at a spread constant of 8.0. The results obtained at this stage were used in the predicting the substituent on the eudesmane skeletons for the test compounds.

Since compounds used in the prediction of the ^{13}C chemical shift values for each of the positions on the Eudesmane skeletons possess some groups that are foreign to the dataset used in [8], the original dataset used in training was expanded from 291 to 377 (by adding the skeletal data of 86 compounds used in the current to the 291 used in previous study).

Additional codes were also introduced to accommodate the new substituent types discovered in the compounds adopted for the first portion of the current study (Appendix B). The scatter plots of codes of substituents against chemical shifts were re-plotted.

In predicting the substituents on the thirty-three (33) test compounds using GRNN, the expanded dataset (^{13}C skeletal chemical shift values of the 377 compounds) were used as input to the system. The corresponding substituent codes at the different position on the eudesmane skeletons (of the 377 compounds) were used as the target data. The ^{13}C values predicted for each position on the eudesmane skeleton of the 33 test compound (at the first stage of this study) were simulated at different spread constant values. The best results were observed at a spread constant of 7.0. Furthermore, the results obtained using GRNN were compared with those got using the chemical shift ranges for each substituent type generated using scatter plots. In the Scatter plots method, the most likely substituent type (over the possible Carbon ranges for each of the 15 positions on the Eudesmane skeleton) was selected. Finally, the percentage recognition of each of the compounds using these procedures were determined from the number of correctly predicted points relative to the total number of positions on each compound.

Results and Discussion

A major key to a successful utilization of the previously described procedures [8, 9] (Alawode and Alawode, 2014 ; Alawode and Alawode, 2015) in achieving a complete structural elucidation is to be able to separate ^{13}C data due to the substituents from those due to the skeleton.

It is also necessary to identify specifically the ^{13}C due to each position on the skeleton. A cursory glance at the 3D plots of whole eudesmanes and their skeletons (Fig. 3a and 3b) shows that the ^{13}C values for whole eudesmane compounds and the skeletal ^{13}C values fall within the same general ranges (0-50, 50-100, 100-150, 150-200 and 200-250) making it difficult to satisfy these conditions. Owing to the complexity of the problem, it was subjected to Artificial Neural Network procedures. ANNs are employed in pattern recognition problems, especially those associated with prediction, classification or control. In the current study, GRNN, an architecture of ANN was utilized in the prediction of ^{13}C values on the different positions of the eudesmane skeleton. Compared to other ANN models, the GRNN is able to converge to the underlying function of the data with only few training samples available [19] (Sun *et al.*,2008).

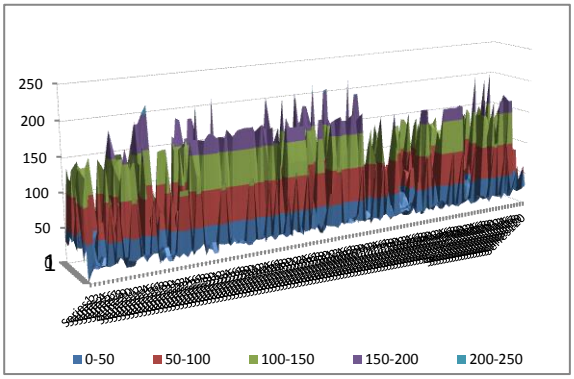

Fig 3a. 3D plot of 13C skeletal data.

Table 1. ^{13}C Chemical Shift Values of Test Compounds.

1	2	3	4	5	6	7	8	9	10
10.9	11.6	12.4	17.7	14.7	12.1	12.2	14.9	13.4	20.7
12.5	12.5	19.7	22.4	20.6	14.4	14.5	20.7	14.1	21.4
23.3	23.0	22.6	25.1	20.6	19.5	15.9	20.8	16.1	21.8
25.3	31.2	23.6	28.8	21.2	20.8	24.4	21.3	18.6	25.5
39.3	33.5	34.2	29.3	21.8	26.7	27.8	24.7	20.6	26.7
41.2	36.0	40.4	31.9	24.6	33.2	38.1	28.7	24.5	30.3
41.7	41.2	46.2	39.2	28.5	41.8	41.9	28.8	29.5	44.1
54.0	42.8	52.3	48.9	28.7	42.8	43.5	39.1	38.2	50.1
81.5	52.3	54.6	57.7	39.1	51.8	53.2	41.4	40.1	54.1
125.9	52.5	69.9	72.0	41.3	66.5	57.1	41.6	40.3	67.7
128.4	78.2	79.4	73.4	41.4	77.7	66.2	49.9	43.6	69.2
151.5	79.3	125.3	75.4	49.5	77.9	71.3	55.3	59.4	70.2
155.1	110.3	162.0	143.2	55.1	120.9	76.5	69.8	59.8	72.3
177.4	142.8	178.2	145.4	69.7	126.8	78.5	71.2	68.2	72.6
186.0	179.4	201.2	201.3	71.1	128.1	120.6	80.4	71.2	77.3
0	0	0	0	80.2	128.5	128.0	118.3	71.4	84.8
0	0	0	0	172.3	134.0	133.2	128.3	71.8	91.7
0	0	0	0	0	138.2	138.5	128.3	75.0	128.2
0	0	0	0	0	167.1	167.1	128.9	125.0	128.2
0	0	0	0	0	169.4	168.9	128.9	125.8	128.3
0	0	0	0	0	0	0	130.6	134.6	128.3
0	0	0	0	0	0	0	134.3	142.6	129.3
0	0	0	0	0	0	0	145.3	166.5	129.4
0	0	0	0	0	0	0	168.2	168.2	129.6
0	0	0	0	0	0	0	0	168.6	129.6
0	0	0	0	0	0	0	0	0	130.2
0	0	0	0	0	0	0	0	0	130.2
0	0	0	0	0	0	0	0	0	133.1
0	0	0	0	0	0	0	0	0	133.4
0	0	0	0	0	0	0	0	0	164.8
0	0	0	0	0	0	0	0	0	165.7
0	0	0	0	0	0	0	0	0	170.0
					0	0	0	0	170.0

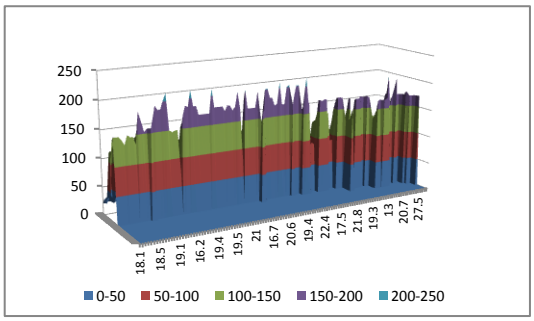

Fig 3b. 3D plot of 13C Full Eudesmane Compounds.

Furthermore, since the task of determining the best values for the several network parameters is difficult and often involves some trial and error methods, GRNN models require only one parameter (the spread constant) to be adjusted experimentally. This makes GRNN a very useful tool to perform predictions and comparisons of system performance in practice. Previous works relating the predictive capability of GRNN to backpropagation neural network and other nonlinear regression techniques highlighted the advantages of GRNN to include excellent approximation ability, fast training time, and exceptional stability during the prediction stage [20, 21](Mahesh *et al.*, 2014; Schneider and Wrede, 1998). GRNN was able to predict the ^{13}C chemical shift values for the different positions with sufficient accuracy with values falling within ranges already determined from the previous study using scatter plots [8, 9] (Alawode and Alawode, 2015). Since neural networks generally learn by examples, the quality or accuracy of their predictions will increase with increase in representation of the substitution patterns of the test compounds in the ^{13}C chemical shift values used for the training of the network. Predictions generally would likely improve as ^{13}C data available for training increase (as substitution patterns in the compounds would certainly become more diverse). Least deviation of ^{13}C values from those of the test compounds was obtained at a spread constant of 8.0. The actual and predicted ^{13}C chemical shift values for each of the fifteen (15) carbon positions on the Eudesmane skeleton for ten (10) of the test compounds are presented in Table 2. The ^{13}C values for all the carbon atoms (comprising both the skeletal and substituent carbon atoms) in each of the ten compounds have been presented in Table 1.

In using the Scatter plots and GRNN approaches to predict whether or not substituents are attached to each position of the eudesmane skeleton (and the nature of the substituents), the original dataset (used for the previous study) was expanded. The new substituent types were coded as described previously [9] (Alawode and Alawode, 2015). For the scatter plots approach, the graphs of codes of substituents against ^{13}C chemical shift values were re-plotted for each position (plots not shown). Significant changes in the pattern of the plots (previously published in [9]Alawode and Alawode (2015)) were observed at C_5, C_{12} and C_{15}. The new ranges were taken into consideration in predicting the substituents on each position. Specifically, the range within which the predicted ^{13}C chemical shifts values falls for each position were identified and the most probable substituent for that range selected for the position.

Table 2. Actual (A) ^{13}C Data of Test Compounds Versus the Predicted (P) Values.

	1		2		3		4		5	
	(A)	(P)	(A)	(P)	(A)	(P)	(A)	(P)	(A)	(P)
C1	155.1	155.0	78.2	77.9	201.2	201.9	31.9	32	80.2	77.5
C2	125.9	126.0	31.2	31.3	125.3	125.1	25.1	25.8	28.5	28.3
C3	186.0	186.1	33.5	34.2	162	151.6	75.4	73.3	41.3	40.0
C4	128.4	128.8	142.8	143.7	69.9	69.5	73.4	72.1	71.1	71.6
C5	151.5	151.9	52.5	52.7	54.5	53.7	48.9	49.0	55.1	53.0
C6	81.5	80.8	79.3	77.4	79.4	79.3	143.2	143.5	69.7	68.0
C7	54.0	49.5	52.3	50.0	52.3	52.2	145.4	144.8	49.5	49.7
C8	23.3	20.3	23.0	25.2	22.6	22.6	201.3	201.5	20.6	24.6
C9	39.3	38.2	36.0	36.1	34.2	33.8	57.7	57.7	41.4	39.7
C10	41.7	41.5	42.8	41.9	46.2	46.2	39.2	39.2	39.1	40.1
C11	41.2	38.2	41.2	41.1	40.4	40.3	72.0	72.0	28.7	32.4
C12	12.5	9.9	12.5	14.0	12.4	12.4	29.3	29.3	20.6	22.0
C13	177.4	178.3	179.4	175.7	178.2	178.4	28.8	28.8	21.2	20.9
C14	25.3	25.2	11.6	11.8	23.6	23.0	17.7	17.7	14.7	13.9
C15	10.9	11.0	110.3	109.6	19.7	22.2	22.4	22.3	24.6	20.5

Table 2. (continues). Actual (A) 13C Data of Test Compounds Versus the Predicted (P) Values.

6		7		8		9		10	
(A)	(P)	(A)	(P)	(A)	(P)	(A)	(P)	(A)	(P)
77.7	75.6	76.8	75.3	80.4	84.5	71.4	71.7	72.6	76.5
26.7	32.6	27.8	30.7	28.7	23.4	24.5	24.7	67.7	200.5
33.2	118.3	38.1	33.3	41.6	42.7	38.2	38.4	44.1	54.7
128.5	133.3	71.3	142.0	71.2	82.4	71.2	71.2	70.2	74.1
126.8	51.3	57.1	52.0	55.3	57.2	43.6	43.6	91.7	91.3
77.9	77.4	78.5	78.4	69.8	69.5	75.0	75.4	69.2	69.4
51.8	53.5	53.2	53.3	49.9	49.9	40.1	40.6	54.1	53.9
66.5	66.0	66.2	66.0	20.8	23.2	68.2	68.3	77.3	77.1
42.8	39.2	43.5	40.3	41.4	33.0	71.8	72.4	72.3	72.1
41.8	40.7	41.9	42.6	39.1	48.3	40.3	40.7	50.1	51.9
134.0	134.2	133.2	134.4	28.8	29.3	134.6	135	84.8	85.8
120.9	119.2	120.5	119.4	21.3	21.6	125	124.7	26.7	26.5
169.4	169.9	168.9	169.9	20.7	20.6	168.6	168.7	30.3	30.3
19.5	12.7	15.9	13.5	14.9	17.3	14.1	14.7	20.7	19.9
20.8	26.4	24.4	110.6	24.7	22.5	29.5	29.7	25.5	25.0

Table 3. Comparison of % Recognition (A and B) of Test Compounds Using GRNN (GR) and Scatter Plot (SP) Methods.

	1			2			3		
POSITION	TEST	GR	SP	TEST	GR	SP	TEST	GR	SP
C1	Δ^1	Δ^1	Δ^1	β-OH	β-OH	β-OH	Oxo	Oxo	**
C2	-	-	-	-	-	-	Δ^2	Δ^2	-
C3	Oxo	Oxo	Oxo	-	-	-	-	-	Δ^3
C4	Δ^4	Δ^4	Δ^4	$\Delta^{4(15)}$	$\Delta^{4(15)}$	Δ^4	α-OH	β-OH	β-OH
C5	-	-	-	-	-	-	-	ONic	-
C6	α-Oxy	α-Oxy	α-OAc	α-Oxy	α-Oxy	α-OAc	α-Oxy	α-Oxy	α-OAc
C7	-	-	-	-	-	-	-	-	-
C8	-	-	-	-	-	-	-	-	-
C9	-	-	-	-	-	-	-	-	-
C10	-	-	-	-	-	-	-	-	-
C11	β	β	β	β	β	β	β	-	β
C12	Oxo,6 αOxy	Oxo,6 αOxy	**	Oxo,6αOxy	$OFuc(OAc)_3$	Oxo,OMe	Oxo,6αOxy	Oxo,6αOxy	**
C13	α	β	**	α	O-trans-Cou	-	α	α	**
C14	β	β	β	β	β	β	β	β	β
C15	-	-	β	-	-	**	β	α	β
A (%)		93.33	73.33		86.67	66.67		73.33	53.33
B (%)		100.00	80.00		86.67	73.33		93.33	60.00

A-With Stereochemistry, B-Without Stereochemistry, **outside all possible ranges for the position

Table 3. Comparison of % Recognition (A and B) of Test Compounds Using GRNN (GR) and Scatter Plot (SP) Methods.

	4			5			6		
POSITION	TEST	GR	SP	TEST	GR	SP	TEST	GR	SP
C1	-	-	-	OH, α-H	OCin	β-OH	β-OH	β-OH	β-OH
C2	-	-	-	-	-	-	-	-	-
C3	α-OH	α-OiBu	β-OAng	-	-	-	-	Δ^3	Δ^3
C4	α-OH	OCin	β-OH	α-OH	β-OH	β-OH	Δ^4	-	Δ^4
C5	α-H	β-Oxy	-	-	-	-	-	α-H	-
C6	Δ^6	Δ^6	Δ^6	β-OAc	β-OAc	β-OAc	6αOxy	αOxy	α-OAc
C7	-	-	-	-	-	-	-	-	-
C8	Oxo	Oxo	Oxo	α	-	-	β-OAng	OAng	α-OAc/β-OBzt
C9	-	-	-	-	-	-	-	-	-
C10	-	-	-	-	-	-	-	-	-
C11	OH	OH	Oxy,α	β	α	β	Δ^{11}	Δ^{11}	Δ^{11},β
C12	-	-	-	-	-	-	Oxo, 6αOxy	Oxo, 6αOxy	Oxo,OMe
C13	-	-	-	-	-	-	-	-	-
C14	β	β	β	β	-	β	β	β	β
C15	β	β	β	β	α	β	-	-	β
A (%)		80.00	73.33		60.00	80.00		73.33	60.00
B (%)		80.00	86.67		93.33	100		80.00	73.33

A-With Stereochemistry, B-Without Stereochemistry, **outside all possible ranges for the position

Table 3. Comparison of % Recognition (A and B) of Test Compounds Using GRNN (GR) and Scatter Plot (SP) Methods.

	7			8			9		
POSITION	TEST	GR	SP	TEST	GR	SP	TEST	GR	SP
C1	β-OH	β-OH	β-OH	β-OH	β-O(α-OH-iVa)	β-OH	β-OH	β-OH	β-OH
C2	-	-	-	-	-	-	-	-	-
C3	-	-	-	-	-	-	-	-	-
C4	α-OH	$\Delta^{4(15)}$	Δ^4	β-OH	β-O(α-OH-iVa)	β-OH	β-OH	β-OH	β-OH
C5	-	α-H	-	-	-	-	-	-	-
C6	6αOxy	αOxy	α-OAc	β-OCin	O2MeBu-(2'OAc,3'OH)	α-OAc	β-Oxy	β-Oxy	α OAc
C7	-	-	-	-	-	-	-	-	-
C8	β-OAng	β-OAng	α-OAc/β-OBzt	-	-	-	β-OEpang	β-OEpang	α OAc/ β-OBzt
C9	-	-	-	-	-	-	α-OAng	α-OAng	β-OBzt
C10	-	-	-	-	-	-	-	-	-
C11	Δ^{11}	Δ^{11}	Δ^{11},β	β	α	β	Δ^{11}	Δ^{11}	Δ^{11},β
C12	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, OMe	-	-	-	Oxo, 6 βOxy	Oxo, 6 βOxy	Oxo, OMe
C13	-	-	-	-	-	-	-	-	-
C14	β	β	β	β	β	β	β	β	-
C15	β	-	**	α	β	β	α	α	-
A (%)		80.00	60.00		66.67	86.67		100.00	53.33
B (%)		93.33	66.67		80.00	93.33		100.00	66.67

A-With Stereochemistry, B-Without Stereochemistry, **outside all possible ranges for the position

Table 3. Comparison of % Recognition (A and B) of Test Compounds Using GRNN (GR) and Scatter Plot (SP) Methods.

	10		
POSITION	TEST	GR	SP
C1	β-OAc	β-OAc	β-OH
C2	β-OH	Oxo	Oxo
C3	-	-	-
C4	α-OH	α-OH	β-OH
C5	αOxy, 11α	αOxy	αOxy
C6	α OAc	α OAc	α OAc
C7	-	-	-
C8	α-OBzt	α-OBzt	α OAc/ β-OBzt
C9	α-OBzt	α-OBzt	β-OBzt
C10	-	-	-
C11	α	βOxy, βOH	**
C12	β	α	-
C13	α	α	-
C14	β	β	β
C15	β	β	β
A (%)		73.33	40.00
B (%)		80.00	86.67

A-With Stereochemistry, B-Without Stereochemistry, **outside all possible ranges for the position.

The scatter plot approach gives all the possible substituents for each position with their corresponding likelihood of occupying the position. The most likely substituent has been selected in this study for the sake of simplicity and to create a basis for comparison with the GRNN method. The substituents predicted by both methods for the ten (10) compounds are shown in Table 3. It is quite possible that the actual substituent for a particular position (within a specified range on the table) has a lesser likelihood of occupying the position. This may affect the accuracy of prediction given by the scatter plot method in this study. Also, a study of the table indicated positions on the eudesmane skeleton of some of the compounds under study where predicted values fall outside all the possible ranges for such positions as shown on the Table (indicated by **). This is due to the fact that the chemical shift ranges for each position were obtained from a plot of codes of

substituents against chemical shift values for specific eudesmane compounds while the values being compared with the ranges are predicted values. Unlike the scatter plot method, GRNN predicted substituents for all the 15 position for all the test compounds. The degree of recognition of the test compounds (from both methods) ranged between 40 and 100%. Again, as indicated earlier, the quality of prediction would likely increase as more training data become available. Percentage recognition generally increased when the stereochemistry (α or β) of the substituents were not considered. The GRNN method showed a slightly better result than the scatter plot method (subject to the limitation that the most likely substituent for the specified range for each position is the actual substituent).

This procedure may be very useful in elucidating structures of unknown eudesmane compounds.

References

[1] F. C. Oliveira, M. J. P. Ferreira, C. V. Nunez, G. V. Rodriguez, V. P. Emerenciano, "^{13}C NMR spectroscopy of eudesmane sesquiterpenes," Progress in Nuclear Magnetic Resonance Spectroscopy, vol. 37: 1–45, 2000.

[2] M. J. P. Ferreira, F. C. Oliveira, G. V. Rodrigues and V. P. Emerenciano, "^{13}C NMR Pattern Recognition of Guaiane Sesquiterpenes," Internet Electronic Journal of Molecular Design, vol. 3(11): 737-749, 2004

[3] V. P. Emerenciano, A. C. Bussolini, G.V. Rodrigues, M. Furlan and D. L. G. Fromanteau, "A pattern recognition method on 13C NMR spectroscopy of sesquiterpene skeletal types," Spectroscopy, vol. 11: 95–115, 1993.

[4] F. C. Oliveira, M. J. P. Ferreira, C. V. Nunez, G. V Rodriguez, V. P. Emerenciano . 13C NMR spectroscopy of Eudesmane Sesquiterpenes. Progress in Nuclear Magnetic Resonance Spectroscopy, vol. 37: 1-45, 2000.

[5] G. V. Rodrigues, I. P. A. Campos, and V. P. Emerenciano, "Applications of artificial intelligence to structure determination of organic compounds **. Determination of groups attached to skeleton of natural products using ^{13} C Nuclear Magnetic Resonance spectroscopy," Spectroscopy, pp. 191-200, 1997.

[6] S. A. V. Alvarenga, J. P. Gastmans, G. V. Rodrigues and V.P. Emerenciano, "Ditregra – an auxiliary program for structural determination of diterpens," Spectroscopy, vol. 13: 227–249, 1997.

[7] P. A. T. Macari, J. P. Gastmans, G. .V. Rodrigues and V. P. Emerenciano, "An expert system for structure elucidation of triterpenes," Spectroscopy, vol. 12:139–166, 1994.

[8]T. T. Alawode and K. O. Alawode, "Prediction of substituent types and positions on skeelton of Eudesmane-type Sesquiterpenes Using Generalized Regression Neural Network (GRNN)," African Journal of Pure and Applied Chemistry, vol. 8(7), pp. 102-109, 2014.

[9] T. T. Alawode and K. O. Alawode, "Validation of Structures of Novel Eudesmane Sesquiterpenes Using Scatter Plots," British Journal of Applied Science and Technology, vol. 7(1): 97-113, 2015.

[10] Q. Wu, Y. Shi and Z. Jia, "Eudesmane Sesquiterpenoids from the Asteraceae Family," Nat. Prod. Rep., vol. 23: 699-734, 2006.

[11] M. T. Scotti, V. Emerenciano, M.. J. P. Ferreira, L. Scotti, R. Stefani, M. .S. da Silva and F. J. B. Mendonça Junior, "Self-Organizing Maps of Molecular Descriptors for Sesquiterpene Lactones and Their Application to the Chemotaxonomy of the Asteraceae Family," Molecules, vol. 17: 4684-4702, 2012

[12] S. A., Hannan, R. R. Manza and R. .J. Ramteke, "Generalized regression neural network and radial basis function for heart disease diagnosis," International Journal of Computer Applications, vol. 7(13):7-13, 2010

[13] Specht, "A general regression neural network," IEEE Transactions on Neural Networks, vol. 2(6):568-576, 1999.

[14] J. Aires-de-Sousa, M. Hemmer and J. Gasteiger, "Prediction of ^{1}H NMR Chemical Shifts Using Neural Networks," Analytical Chemistry, vol. *74*(1), pp. 80-90, 2002.

[15] Y. Binev and J. Aires-de-Sousa, "Structure-Based Predictions of ^{1}H NMR Chemical Shifts Using Feed-Forward Neural Networks," Chem. Inf. Comput. Sci., vol. 44: 940-945, 2004.

[16] L. Fraser and D. A. Mulholland, "A robust technique for group classification of the C-13 NMR spectra of natural products from *Meliaceae* Fresenius," J Anal Chem., vol. 365:631-634, 1999.

[17] A. Rahman and V. U. Ahmad, ^{13}C-NMR of Natural Products. Volume 1. Monoterpenes and Sesquiterpenes. Springer Science+Business Media New York, 1992, pp. 606-683.

[18] The MathWorks, Inc. "MATLAB and Statistics Toolbox Release", Natick, Massachusetts, United States, 2009a

[19] G. Sun, S. J. Hoff, B. C. Zelle, and M. .A. Nelson, M, "Development and comparison of backpropagation and generalized regression neural network models to predict diurnal

and seasonal gas and pm10 Concentrations and emissions from swine buildings," American Society of Agricultural and Biological Engineers, vol. 51(2):685-694, 2008.
[20] C. Mahesh, E. Kannan and M. S. Saravanan, "Generalized regression neural network based expert system for hepatitis b diagnosis," Journal of Computer Science, vol. 10(4):563-569, 2014.
[21] G. Schneider and P. Wrede, "Artificial Neural Networks for Computer-based Molecular Design," Progress in Biophysics & Molecular Biology, vol. 70:175-222, 1998.

Appendix A: List of Test Compounds

Code	Compound
1	α-Santonin
2	11β,13-Dihydroreynosin
3	Vulgarin
4	Plucheinol
5	1β, 4α-dihydroxy-6β-acetoxy eudesmane
6	1β-Hydroxy-8β-Tigloxy-4(11)(13)-eudesmadien-6α, 12-olide
7	1β, 4α-Dihydroxy-8β-Tigloxy-11(13)-Eudesmen-6α, 12-olide
8	6β-Cinnamoyloxy-1β,4β-Dihydroxy eudesmane
9	9α-Angeloxycalostephanolide-8-O-[2S, 3S-Epoxy-2-methylbutyrate]
10	Triptofordin C-2

Appendix B: Newly Encountered Substituents types and their assigned Codes

Substituent	Codes	Substituent	Codes
Oxo,8Oxy	216	11β, 6 αOxy	246
Oxo,8αOxy	217	11β, 6 βOxy	247
Oxo, 8β-Oxy	219	αH, OAc	248
Oxo, 6Oxy	220	11β, 11βOxy, 12α,12βOxy	252
Oxo,6αOxy	221	1αOxy, 4αOxy	253
Oxo, 6β-Oxy	222	OBz, 4'-OMe	254
15-Oxy, 6Oxy	223	αOBz, 4'-OMe	255
15-Oxy, 6αOxy	224	βOBz, 4'-OMe	256
15-Oxy, 6 βOxy	225	OCinn, 4'OH	257
15 αOxy, 6Oxy	226	αOCinn, 4'OH	258
15 βOxy, 6Oxy	227	βOCinn, 4'OH	259
15 αOxy, 6αOxy	228	OEt	260
15 βOxy, 6 βOxy	229	αOEt	261
1αOxy, 2αOxy	230	βOEt	262
3 αOxy, 4αOxy	231	11Oxy, 5Oxy	263
11-NCS, 11α	232	11αOxy, 5αOxy	264
11-NCS, 11β	233	11αOxy, 5βOxy	265
αOxy, αOH	234	11 βOxy, 5αOxy	266
αOxy, βOH	235	11 βOxy, 5αOxy	267
βOxy, αOH	236	αH, αOH	268
βOxy, βOH	237	αH, βOH	269
4-Oxy,5-Oxy	238	βH, αOH	270
5-Oxy,6-Oxy	239	βH, βOH	271
8-Oxy, 12-Oxy	240	Δ^8, 8Oxy	272
11α, 6αOxy	244	Δ^8, 8αOxy	273
11α, 6βOxy	245	Δ^8, 8βOxy	274