STRUCTURAL ELUCIDATION OF GUAIANE SESQUITERPENES FROM ^{13}C DATA USING GENERALIZED REGRESSION NEURAL NETWORK (GRNN) AND SCATTER PLOTS METHODS

Taye T. Alawode [1], Kehinde O. 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**

Corresponding Author: Taye T. Alawode, onatop2003@yahoo.com

ABSTRACT: This study seeks to elucidate of structures of unknown Guaiane sesquiterpenes from their ^{13}C chemical shift values using Generalized Regression Neural Network (GRNN) and scatter plot methods. The ^{13}C values for each of the fifteen (15) positions of the skeletons of the Guaiane sesquiterpenes were predicted using GRNN. From these predicted values, the substituents attached to each of the positions were predicted using GRNN and Scatter Plot methods. In predicting the skeletal ^{13}C values, the ^{13}C data of 116 Guaiane sesquiterpenes were used as input to GRNN while their corresponding data were used as the target data. The network was trained and simulated using twenty-five (25) test compounds. The best results were obtained at a spread constant of 5.0. In predicting the substituents on each position on the skeleton using the GRNN approach, the predicted ^{13}C skeletal values of the test compounds were simulated following training of the GRNN using skeletal data of the 116 compounds as input data and their corresponding substituents (given codes) as the target data. The best results were obtained at a spread constant of 5.0. In the Scatter Plots method, graphs of codes of substituents for the 116 guaiane sesquiterpenes were plotted against the corresponding ^{13}C chemical shift values of the skeletal Carbon to which they were attached. This gave the range of values over which each substituent may be obtained. The most likely substituent(s) for each position were selected. The degree of recognition of the test compounds (from both methods) ranged between 46.67 and 100%. Both methods gave similar recognition rates for the test compounds. GRNN and Scatter plots demonstrated great potential for use in the structural elucidation of unknown compounds from ^{13}C values.

KEYWORDS*:* GRNN, Scatter plots, structural elucidation, ^{13}C, Guaiane sesquiterpenes

1. INTRODUCTION

Characterization of structures of bioactive compounds isolated from plants and other sources is central to the success of any research in natural products chemistry. To this end, several research efforts have been directed at simplifying the process of structures of compounds isolated from nature (utilizing expert systems) have been described several authors including Emerenciano et al. [E+93], Oliveira et al. [O+00], Alvarenga et al. [A+97], and Macari [M+94]. The current study explores GRNN (an architecture of Artificial Neural Networks) and Scatter plot procedures in elucidating the structures of Guaiane sesquiterpenes. The Guaiane skeleton is shown in Figure 1.

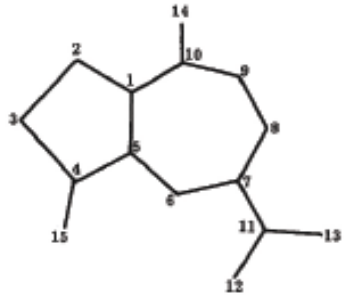


Fig.1: Guaiane skeleton

Typical substituents found in sesquiterpenes have been published by Wu et al. [WSH06] and Rodrigues et al.[RCE97]. The theories of ANNs and GRNN and their applications have been discussed by several authors including Scotti et al. [S+12], Hannan et al. [HMR10], Specht [Spe99], Aires-de-Sousa et al.[AHG02], Binev and Aires-de-Sousa[BA04] and Fraser and Mulholland [FM99].

2. METHODOLOGY

A total of 141 Guaiane-type sesquiterpenes were used in the current study. Of these total, twenty-five compounds were used as the test data. Owing to space limitations, the ^{13}C chemical shift data of ten of the test compounds (arranged in ascending order) are given in Table 1. The names of the compounds are provided in Appendix A. The remaining compounds were used in training the network. The ^{13}C spectral data used were extracted from structures of Guaiane sesquiterpenes published by Atta-ar-Rahaman and Ahmad [RA99]. In order to achieve a complete structural elucidation of the test compounds, the study

was carried out in two stages. The first stage involves separating the ^{13}C data due to the substituents from those due to the skeleton and identifying specifically the ^{13}C due to each position on the skeleton. To solve this problem, GRNN was employed. The ^{13}C data of the training set were entered in ascending order in an Excel sheets. In order to ensure equal number of entry for all the compounds, the difference in ^{13}C data was made up with zeroes for deficient compounds. The corresponding skeletal data for each of the 15 carbon position of the each of the compounds were entered on a second excel sheet and used as the target data. A third excel sheet prepared like the first sheet contains the ^{13}C data for the test compounds. The data were transferred into the Neural Network toolbox of MATLAB 7.8.0 [Mat09a]. 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. Generalized Regression 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 5.0. The predicted skeletal data for each of the test compounds were used as the test data for the second phase of the work.

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

1	2	3	4	5	6	7	8	9	10
20.6	35.9	13.0	13.3	17.4	12.1	17.4	12.1	14.2	12.2
21.9	39.0	24.2	18.7	36.9	12.7	21.0	25.5	14.5	34.7
22.2	40.0	27.5	19.3	38.8	14.4	39.3	28.9	17.4	41.2
22.7	41.3	29.5	22.8	45.8	14.8	48.3	32.1	36.1	41.2
22.9	49.3	36.2	40.6	47.6	20.2	53.8	37.2	47.6	45.2
28.5	72.7	42.8	52.6	48.5	26.0	56.4	44.0	52.4	47.5
39.7	73.1	48.6	54.5	52.7	36.7	58.2	50.4	55.3	54.3
52.0	85.0	52.8	59.0	53.0	37.0	59.6	52.6	56.0	62.6
55.9	108.0	71.6	59.4	54.2	41.1	69.0	62.6	56.3	71.6
59.8	112.0	77.1	65.4	69.2	51.0	79.1	71.5	56.4	74.8
72.7	119.1	78.4	78.4	75.1	77.6	81.3	73.2	57.3	75.0
79.9	141.0	142.3	120.0	75.5	79.7	119.4	75.0	67.4	78.6
134.6	153.5	168.6	132.4	77.4	84.9	122.2	78.5	76.3	78.7
142.1	155.3	176.6	134.2	117.9	128.0	127.4	78.7	81.1	81.2
202.9	170.2	205.5	135.5	122.0	130.9	129.9	85.5	122.5	102.1
0	0	0	145.3	138.2	138.1	134.9	101.8	127.5	112.9
0	0	0	167.7	143.1	155.3	142.1	113.6	129.0	116.5
0	0	0	168.7	169.2	167.2	147.3	118.5	131.6	116.5
0	0	0	169.0	170.4	179.5	147.6	142.0	134.2	117.3
0	0	0	194.7	0	197.5	165.5	151.4	141.8	121.8
0	0	0	0	0	0	169.6	170.1	145.4	125.6
0	0	0	0	0	0	171.4	0	149.5	131.0
0	0	0	0	0	0	0	0	165.3	131.0
0	0	0	0	0	0	0	0	166.6	138.2
0	0	0	0	0	0	0	0	169.5	143.3
0	0	0	0	0	0	0	0	0	151.4
0	0	0	0	0	0	0	0	0	158.3
0	0	0	0	0	0	0	0	0	169.2
0	0	0	0	0	0	0	0	0	171.2

Having predicted the skeletal data for each of the fifteen (15) Carbon positions on the Guaiane skeletons of the test compounds, the second stage of the work involving the prediction of the substituents attached to each of these positions was carried out. To achieve this, two procedures (GRNN and Scatter Plots) were employed and their degrees of accuracy compared. In the GRNN approach, the target data in the first part of the study was used as the training data. In preparing the target data for the second stage, the substituents for each of the 15-carbon positions (for the 116 compounds used in training) were entered (as codes) in place of the ^{13}C chemical shift values for the compounds. The procedures for converting substituents to codes have been described previously by Alawode and Alawode [AA14]. The test data utilized at this stage are those ^{13}C predicted for each of the 15 positions on the Guaiane skeletons for the test compounds from the first stage. After training, the network was simulated using the test

data. The best results were obtained at a spread constant of 5.0. In the Scatter Plots approach, scatter plots of the codes of the substituents against the corresponding ^{13}C values for each of the 15 Carbon positions were obtained. From the patterns obtained on the graphs, ^{13}C ranges over which each substituent may be observed for each of the carbon positions were determined and a reference table of values generated (Table3). The likelihood (in percentages) that a particular substituent would occupy a particular position within a chemical shift range is given in parenthesis in front of each substituent type. This table was then utilized in assigning possible substituents to each of the positions of the skeletons of the Guaianes tested. While the table lists all the possible substituents for each position on the skeleton within particular ^{13}C chemical shift ranges, the most likely substituents (for simplicity) were assigned for this study. The scatter plot diagrams are shown in Appendix B.

3. RESULTS AND DISCUSSION

In the current study, GRNN, was utilized in the prediction of ^{13}C values on the different positions of the guaiane skeleton. GRNN is a very useful tool to perform predictions and comparisons of system performance in practice especially since the model require only one parameter (the spread constant) to be adjusted experimentally. Also, the model has excellent approximation ability, fast training time, and exceptional stability during the prediction stage (Mahesh et al [MKS14]; Schneider and Wrede, [SW98]). A comparison of the actual ^{13}C skeletal data of the test compounds with the values predicted by GRNN is shown in Table 2. The best results were obtained at a spread constant of 5.0.

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	55.9	53.4	41.3	45.2	142.3	143.7	134.2	132.6	45.8	46.0
C2	22.9	25.2	39.0	39.2	205.5	206.6	194.7	194.9	38.8	38.9
C3	28.5	27.4	73.1	73.7	52.8	51.0	135.5	135.9	75.5	75.5
C4	79.9	82.4	155.3	152.4	77.1	75.4	168.7	169.1	69.2	69.4
C5	52.0	51.7	49.3	51.3	168.6	166.1	54.5	53.0	52.7	53.2
C6	39.7	37.0	85.0	79.0	78.4	76.3	78.4	78.6	77.4	77.5
C7	134.6	134.0	35.9	51.0	48.6	47.8	52.6	55.1	47.6	47.8
C8	202.9	203.0	40.0	71.9	24.2	23.8	65.4	65.6	75.1	75.3
C9	59.8	50.2	72.7	41.3	36.2	35.8	40.6	41.1	36.9	36.4
C10	72.7	73.2	153.5	142.7	71.6	71.7	145.3	145.8	143.1	142.7
C11	142.1	143.7	141.0	138.1	42.8	42.6	132.4	134.3	138.2	138.6
C12	21.9	22.1	170.2	169.9	176.6	176.4	169.0	167.8	169.2	169.1
C13	22.2	22.8	119.1	123.2	13.0	12.8	120.0	120.4	122.0	121.1
C14	20.6	32.3	112.0	117.1	27.5	27.2	22.8	23.1	117.9	118.0
C15	22.7	25.0	108.0	113.2	29.5	26.9	18.7	19.7	48.5	48.6

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

	6		7		8		9		10	
	(A)	(P)	(A)	(P)	(A)	(P)	(A)	(P)	(A)	(P)
C1	155.3	155.1	53.8	50.6	50.4	54.3	55.3	54.6	46.3	45.2
C2	36.7	37.0	79.1	80.2	32.1	30.0	76.3	74.6	38.4	34.7
C3	77.6	77.7	129.9	126.1	151.4	104.8	129.0	128.8	80.2	151.4
C4	79.7	80.0	147.3	147.9	113.6	145.2	149.5	148.8	149.5	112.9
C5	51.0	51.4	56.4	55.8	52.6	53.0	52.4	51.9	52.4	54.3
C6	25.0	25.1	81.3	80.0	85.5	83.6	81.8	82.0	78.6	81.2
C7	41.4	41.4	48.3	47.9	44.0	43.1	47.6	47.2	47.3	47.5
C8	84.9	85.0	69.0	68.3	25.5	24.5	67.4	67.2	74.6	74.8
C9	197.5	197.8	39.3	39.0	37.2	33.3	36.1	36.2	36.9	41.2
C10	130.9	131.5	142.1	138.7	73.2	80.8	56.0	73.6	143.2	143.3
C11	37.0	36.8	134.9	133.7	142.0	142.5	134.2	134.6	139.0	138.2
C12	179.5	179.7	169.6	169.1	170.1	170.3	169.5	169.8	169.1	169.2
C13	12.7	12.7	122.2	122.5	118.5	118.3	122.5	121.8	121.2	121.8
C14	14.8	14.9	119.4	120.1	28.9	28.0	56.3	55.0	117.5	117.3
C15	20.2	20.3	17.4	17.2	12.1	37.4	17.4	17.9	115.1	12.2

Table 3: Chemical Shift Ranges For Substituents on the Guaiane Skeleton

Skeletal Carbon	Chemical Shift Range	Codes of substituents (%)
C1	33.4-65.2	Nil (95.77), αH(1.41), 1αOxy,10αOxy(1.41), C_1-C_6 (1.41)
	74.1-98.7 (isolated)	Nil(14.29), αOOH(14.29), 1αOxy,8αOxy(28.57), 1βOxy,2βOxy(14.29), C_1-O-O-C_4 (14.29), αC_1-O-O-C_4α(14.29)
	125.2-159.4	Δ^1(13.89), $\Delta^{1(10)}$ (72.22), 1αOxy,10αOxy(2.78), $\Delta^{1(5)}$(11.11)
C2	21.4-47.8	Nil (100)
	70.1-80.2	βOH(50), αOH(20), αOAc(10), βOAc(10), βOxy (10),
	86.4	1βOxy,2βOxy (isolated)
	124-133.6	Nil(50), Δ^2(50)
	194.3-209.1	Oxo (95.83), Nil(4.17)
C3	27.4-65.3	Nil(100)
	63.1-87.3	βOH(26.67),αOH(10), OH(3.33),αOAc(3.33), βOAc(10), βOGly(10), Oxo(3.33), αOAng(6.67), β-α-OMeAcr-(4'OH) (3.33), β-OMeAcr(3.33), H,αOH(3.33), 3αOxy,4αOxy(13.33), 3βOxy,4βOxy (3.33) ,
	125.2-147.2	Δ^3(88.37), Nil(4.65), Δ^3,OH(2.32), Δ^3,OGly(4.65)
	206.9-218.9	Oxo (100)
C4	30.1-45.3	Nil(100)
	62.3-69.7	4Oxy,5Oxy(9.09), 4αOxy,5αOxy(18.18), 3αOxy,4αOxy(54.55), 4αOxy,15αOxy(9.09), 15β, 4αOxy,15αOxy(9.09)
	74.2-78.5	Nil(33.33), 3βOxy,4βOxy(16.66), αC_1-O-O-C_4α (50)
	79.5-87.4	βOH(16.67), αOH(58.33), β(16.67), OH(8.33)
	112.3-112.9	Nil (isolated)
	134.5-157.9	Nil(44.44), Δ^4(24.44), $\Delta^{4(15)}$(31.11)
	165.1	Nil (isolated)
	168.3-171.6	Nil (100)
	177.7	Nil (isolated)
	178.6	Nil (isolated)
	180.4 ()	Nil(isolated)
C5	30.7-68.0	Nil (89.06), αH(7.81), 4αOxy,5αOxy(3.13)
	80.8-84.7(isolated)	αOH(3), OH(3), 8βOH,5βOxy,8βOxy(1)
	127.7-175.9	Nil(70), Δ^5(25), Δ^4(5)
C6	10.0-39.9	αOxy(5), Nil(90), C_1-C_6(5)
	68.8-86.0	αOxy (93.33), βOxy(5), βH, αOH(1.67)
	119.5-130.6	Nil(4)
	153.3	Oxy (isolated)
	197	Oxo (isolated)
C7	29.7-75.6	Nil (97.20), αH(1.87), 7Oxy,11αOxy(0.93)
	119.8-150.9	$\Delta^{7(11)}$(85.71), Δ^7(14.29)
C8	23.5-41.0	Nil (100)
	62.4-85.5	Nil(1.49), αOH(14.93), βOH(2.98), αOAc(4.48), βOAc(1.49), βOxy(5.97), αOxy(4.48), Oxy(10.45), αOiBu(1.49), β-OEpang(2.98), OAng(1.49), αOAng(7.46), βOAng(4.48), α-OMeAcr-(4'OH)(1.49), β-OTig(2.98), αOMeAcr(4.48), αEpBu^i(4.48), β-4,5-diOH-Tig(2.98), β5[5-OH-Tig]-Tig(4.48), 1αOxy,8αOxy(2.98), αOAc,βH(1.49), β-4-OH-Sarr(1.49), β-4OH-Tig(1.49), αH,β-4,5-diOH-Tig(1.49), β-2,3-diOH-MeBu(1.49), β-4OH-5ac-Tig(1.49), α-p-OH-phenyl-ac(2.98)
	97.8	8 βOH, 5βOxy,8βOxy(isolated)

	124.8	Δ^{8}(isolated)
	158.3	Oxy(isolated)
	199-204.6	Oxo(100)
C9	28.8-53.9	Nil(98.63), β-OEpang(1.37)
	69.5-73.8	αOH(42.86), βOH(14.29), αOAc(14.29), αOAng(28.57)
	80.6	αOGly (isolated)
	98.4	αOAc (isolated)
	122.6-133.6	$\Delta^{9(10)}$ (87.5), $\Delta^{10(14)}$(12.5)
	143.8	Nil(isolated)
	197.8-198.3	Oxo (100)
C10	27.2-41.4	Nil(100)
	55.6-82.5	OH(20), αOH(36), βOH(12), βOAc(4), αOGly(12), 1αOxy,11αOxy(4), 10αOxy,14αOxy(12)
	93.6-93.7	αOH(100)
	125.3-159	Nil(54.24), $\Delta^{10(14)}$(44.07), $\Delta^{1(10)}$(1.69)
C11	36.8-47.1	Nil(53.33), β(43.33), αMe(3.33)
	55.9	β (isolated)
	61.8	7Oxy,11αOxy (isolated)
	73.9-74.0	Nil(isolated)
	76.8	11β,11αOH (isolated)
	77.8	β, αOAng (isolated)
	78	11β, 11αOAc (isolated)
	85	β
	99.6	Nil (isolated)
	119.2-150.3	Nil(6.25), Δ^{11}(35.94), Δ^{11},β(57.81)
C12	19.0-28.3	Nil(100)
	109.0-113.6	Nil (100)
	138.4	8Oxy,12Oxy (isolated)
	167.5-180.9	OAc(1.61), Oxo,OH(4.84), Oxo,8Oxy (8.06), Oxo,8βOxy (3.22), Oxo,6Oxy(3.22), Oxo,6αOxy(69.35), Oxo,6βOxy(9.68)
C13	8.8-29.3	Nil(41.46), α(29.27), β(29.27)
	63.2-65.9	OH(75), OAc(25)
	116.9-126.5	Nil(100)
C14	12.1-32.3	Nil(43.33), α(23.33), β(33.33)
	54.7-56.2 (isolated)	10αOxy, 14αOxy (60), Cl(20), βCl(20)
	69.5	βOH (isolated)
	71.4	β-5OH-Tig (isolated)
	109.5-120.1	Nil(100)
	169.5 (isolated)	
	170.3(isolated)	
C15	7.6-30.9	Nil(100)
	48.4-69.4	OH(20), OAc (10), OGly(40), 4αOxy,15αOxy(10), 15β,4αOxy,15αOxy(10), βCl(10)
	105.1-121.6	Nil (100)
	186.6	Oxo (isolated)

Table 4 compares the substituent predicted from both procedures (GRNN and Scatter Plots) with the actual substituents on the ten test compounds earlier selected. The percentage recognition of the compounds ranged between 46.67% and 100% for the GRNN and scatter plot methods. Also, for some positions on the guaiane skeleton of some of the test compounds in the Scatter plots approach, the

predicted ^{13}C skeletal values fell outside all the possible ranges for such positions. This may be due to a poor representation of the substitution patterns of the affected test compound in the training set. Generally, recognition rates of the test compounds obtained using both methods are only slightly different from each other. Percentage recognition generally increased slightly when the stereochemistry (α or β) of the substituents were not considered.

Table 4: 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	-	-	-	-	-	-	$\Delta^{1(5)}$	$\Delta^{1(5)}$	$\Delta^{1(10)}$
C2	-	-	-	-	-	-	Oxo	Oxo	Oxo
C3	-	-	-	βOH	OCin	βOH	-	-	-
C4	βOH	βOH	αOH	$\Delta^{4(15)}$	$\Delta^{4(15)}$	-	OH	αOH	$\alpha C_1OOC_4\alpha$
C5	-	-	-	-	-	-	-	-	-
C6	-	-	-	αOxy	αOxy	αOxy	αOxy	αOxy	αOxy
C7	$\Delta^{7(11)}$	$\Delta^{7(11)}$	$\Delta^{7(11)}$	-	-	-	-	-	-
C8	Oxo	Oxo	Oxo	-	OGly-(2', 6'-OAc)	αOH	-	-	-
C9	-	-	-	αOH	-	-	-	-	-
C10	αOH	OH	αOH	$\Delta^{10(14)}$	$\Delta^{10(14)}$	-	OH	OH	αOH
C11	-	-	Δ^{11},β	Δ^{11},β	***	Δ^{11},β	-	-	-
C12	-	-	-	Oxo, 6 αOxy	Oxo, 6 αOxy	Oxo, 6 αOxy	Oxo, 6 αOxy	Oxo, 6 αOxy	Oxo, 6 αOxy
C13	-	-	-	-	-	-	α	α	-
C14	β	α	-	-	-	-	α	α	-
C15	α	α	-	-	-	-	α	-	-
A		86.67	73.33		73.33	73.33		86.67	60.00
B		100.00	93.33		73.33	73.33		100.00	86.67

**outside scope A- With Stereochemistry, B- Without Stereochemistry

Table 4 (continues): Comparison of % Recognition 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	$\Delta^{1(5)}$	$\Delta^{1(5)}$	$\Delta^{1(10)}$	-	-	-	$\Delta^{1(5)}$	$\Delta^{1(5)}$	$\Delta^{1(10)}$
C2	Oxo	Oxo	Oxo	-	-	-	-	-	-
C3	Δ^{3}	Δ^{3}	Δ^{3}	βOH	βOH	βOH	αOTig	β-Nor	βOH
C4	-	-	-	15β,4αOxy,15αOxy	4βOxy,15βOxy	3αOxy, 4αOxy	αOH	αOH	αOH
C5	-	-	-	-	-	-	-	-	-
C6	αOxy	αOxy	αOxy	αOxy	αOxy	αOxy	-	-	-
C7	-	-	-	-	-	-	-	-	-
C8	OEpang	11α, 6βOxy	αOH	αEpBu^{i}	11α, 6βOxy	αOH	Oxy	Oxy	αOH
C9	-	-	-	-	-	-	Oxo	Oxo	Oxo
C10	-	-	-	$\Delta^{10(14)}$	$\Delta^{10(14)}$	-	-	-	-
C11	Δ^{11}	***	Δ^{11},β	Δ^{11},β	Δ^{11},β	Δ^{11},β	-	-	-
C12	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 8Oxy	Oxo, 8Oxy	Oxo, 6αOxy
C13	-	-	-	-	-	-	β	β	-
C14	-	-	-	-	-	-	-	-	-
C15	-	-	-	15β,4αOxy,15αOxy	4βOxy,15βOxy	OGly	β	β	-
A		86.67	80.00		80.00	73.33		93.33	60.00
B		86.67	86.67		93.33	73.33		93.33	73.33

**outside scope of ranges predicted by scatter plots method A- With Stereochemistry, B- Without Stereochemistry

Table 4 (continues): Comparison of % Recognition 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	-	-	-	-	-	-	-	-	-
C2	βOH	βOAc	βOH	-	-	-	βOH	βOH	βOH
C3	Δ^3	Δ^3	Δ^3	Δ^3,OGly	Δ^3	**	Δ^3	Δ^3	Δ^3
C4	-	-	-	-	-	-	-	-	-
C5	-	-	-	-	-	-	-	-	-
C6	αOxy	αOxy	αOxy	αOxy	αOxy	αOxy	αOxy	αOxy	αOxy
C7	-	βOH	-	-	-	-	-	-	-
C8	β-4OH-Tig	β-4,5-diOH-Tig	αOH	-	-	-	β5[5-OH-Tig]-Tig	β5[5-OH-Tig]-Tig	αOH
C9	-	-	-	-	-	-	-	-	-
C10	$\Delta^{10(14)}$	$\Delta^{10(14)}$	-	αOH	α-OGly	αOH	10αOxy, 14αOxy	αOH	αOH
C11	Δ^{11},β	Δ^{11}	Δ^{11},β	Δ^{11},β	Δ^{11},β	Δ^{11},β	Δ^{11}	Δ^{11},β	Δ^{11},β
C12	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy
C13	-	-	-	-	-	-	-	-	-
C14	-	-	-	β	β	-	10αOxy, 14αOxy	βCl	10αOxy, 14αOxy
C15	-	-	-	-	-	**	-	-	-
A		73.33	86.67		86.67	80.00		80.00	80.00
B		80.00	86.67		86.67	86.67		86.67	86.67

**outside scope of ranges predicted by scatter plots method A- With Stereochemistry, B- Without Stereochemistry

Table 4 (continues): Comparison of % Recognition of Test Compounds Using GRNN (GR) and Scatter Plot (SP) Methods

	10		
POSITION	**TEST**	**GR**	**SP**
C1	-	-	-
C2	-	-	-
C3	βOGly	Δ^3, OGly	**
C4	$\Delta^{4(15)}$	-	-
C5	-	-	-
C6	αOxy	αOxy	αOxy
C7	-	-	-
C8	α-p-OH-phenyl-ac	βH,OH	αOH
C9	-	-	-
C10	$\Delta^{10(14)}$	$\Delta^{10(14)}$	-
C11	Δ^{11},β	Δ^{11},β	Δ^{11},β
C12	Oxo, 6αOxy	Oxo, 6αOxy	Oxo, 6αOxy
C13	-	-	-
C14	-	-	-
C15	-	-	-
A		80.00	73.33
B		80.00	73.33

**outside scope of ranges predicted by scatter plots method A- With Stereochemistry, B- Without Stereochemistry

CONCLUSION

Generalized Regression Neural Network and Scatter Plots methods have proved to be useful in the elucidation of structures of Guaiane-type sesquiterpenes. The study would be expanded to other classes of natural products in order to establish the suitability of the procedures in determining their structures from ^{13}C NMR chemical shift values.

REFERENCES

[AA14] **T. T. Alawode, 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. 8(7): 102-109, 2014.

[AHG02] **J. Aires-de-Sousa, M. Hemmer, J. Gasteiger** - *Prediction of ^{1}H NMR Chemical Shifts Using Neural Networks*, Analytical Chemistry. 74(1), 80-90, 2002.

[A+97] **S. A. V. Alvarenga, J. P. Gastmans, G. V. Rodrigues, V. P. Emerenciano** - *Ditregra – an auxiliary program for structural determination of diterpens*, Spectroscopy **13:** 227–249, 1997.

[BA04] **Y. Binev, J. Aires-de-Sousa** - *Structure-Based Predictions of ^{1}H NMR Chemical Shifts Using Feed-Forward Neural Networks*, Chem. Inf. Comput. Sci. 44: 940-945, 2004.

[E+93] **V. P. Emerenciano, A. C. Bussolini, G. V. Rodrigues, M. Furlan, D. L. G. Fromanteau** - *A pattern recognition method on 13C NMR spectroscopy of sesquiterpene skeletal types*, Spectroscopy 11: 95–115, 1993.

[FM99] **L. Fraser, D. A. Mulholland** - *A robust technique for group classification of the C-13 NMR spectra of natural products from* Meliaceae. Fresenius J Anal Chem. 365:631-634, 1999.

[HMR10] **S. A. Hannan, R. R. Manza, R. J. Ramteke** - *Generalized regression neural network and radial basis function for heart disease diagnosis*. International Journal of Computer Applications. 2010;7(13):7-13, 2010.

[Mat09] **The MathWorks, Inc** - *MATLAB and Statistics Toolbox Release*. Natick, Massachusetts, United States, 2009.

[MKS14] **C. Mahesh, E. Kannan, M. S. Saravanan** - *Generalized regression neural network based expert system for hepatitis b diagnosis*. Journal of Computer Science. 10(4):563-569, 2014.

[M+94] **P. A. T. Macari, J. P. Gastmans, G. V. Rodrigues, V. P. Emerenciano** - *An expert system for structure elucidation of triterpenes*. Spectroscopy. 12:139–166, 1994.

[O+00] **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. 37: 1–45, 2000.

[RA92] **A. Rahman, V. U. Ahmad** - *^{13}C-NMR of Natural Products. Volume 1. Monoterpenes and Sesquiterpenes*. Springer Science+Business Media New York. pp. 606-683, 1992.

[RCE97] **G. V. Rodrigues, I. P. A. Campos, 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. 191-200, 1997.

[Spe99] **D. F. Specht** - *A general regression neural network.* IEEE Transactions on Neural Networks. 2(6):568-576, 1999.

[SW98] **F. Schneider, P. Wrede** - *Artificial Neural Networks for Computer based Molecular Design*. Progress in Biophysics & Molecular Biology. 70:175-222, 1998.

[S+12] **M. T. Scotti, V. Emerenciano, M. J. P. Ferreira, L. Scotti, R. Stefani, M. S. da Silva, 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. 17: 4684-4702, 2012.

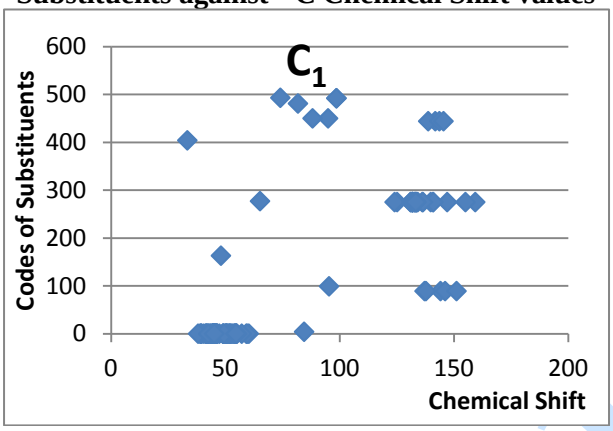
[WSJ06] **Q. Wu, Y. Shi, Z. Jia** - *Eudesmane Sesquiterpenoids from the Asteraceae Family*. Nat. Prod. Rep. 23: 699-734, 2006.

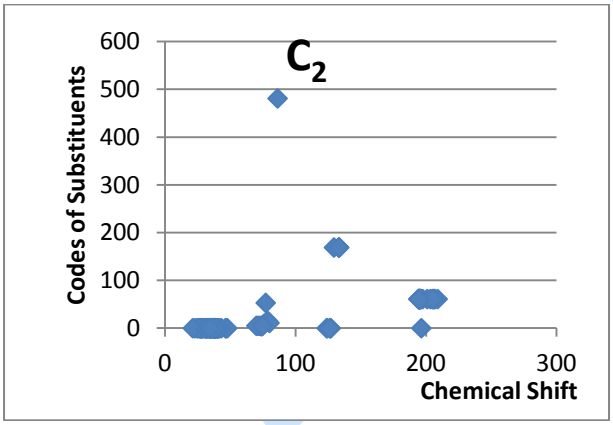
APPENDIX

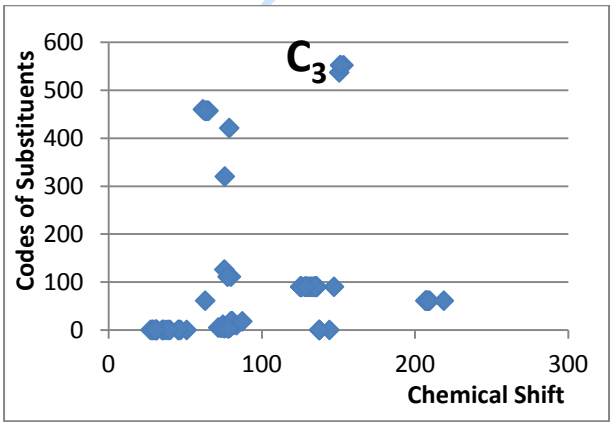
Appendix A. List of Test Compounds

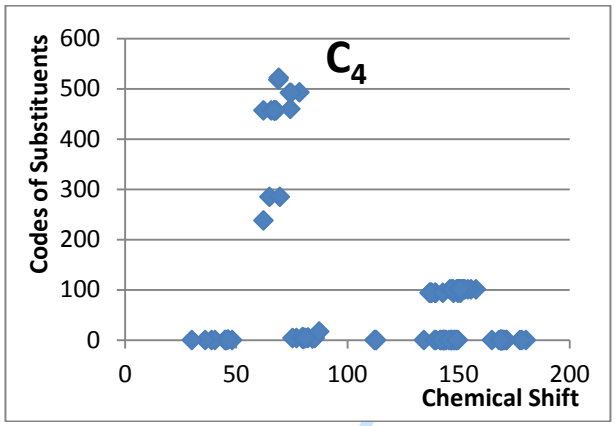
Code	Name
1	Zedoarondiol
2	9α-Hydroxyzaluzanin C
3	Artabsinolide B (10)
4	Sesquiterpene HG-I
5	Subluteolide
6	3-O-Tigloylcarolenalone
7	8β-(4'-Hydroxy-5'-acetoxytigloyloxy)-Preeupatundin
8	Prenantheside A
9	Desacetylspicatin
10	Crepiside F

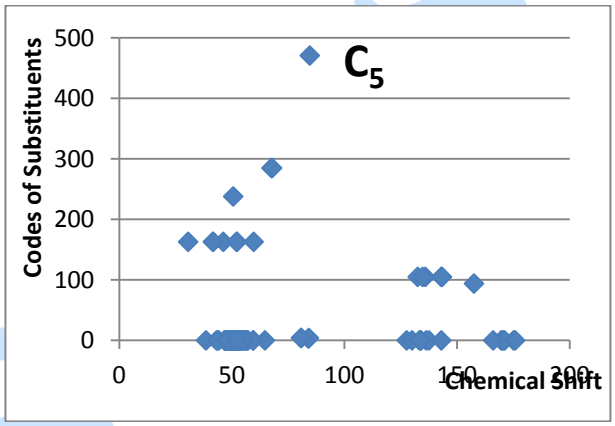
Appendix B. Scatter Plots of Codes of Substituents against ^{13}C Chemical Shift values

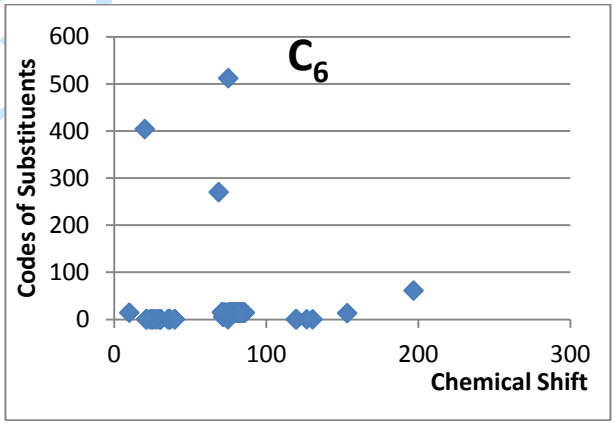


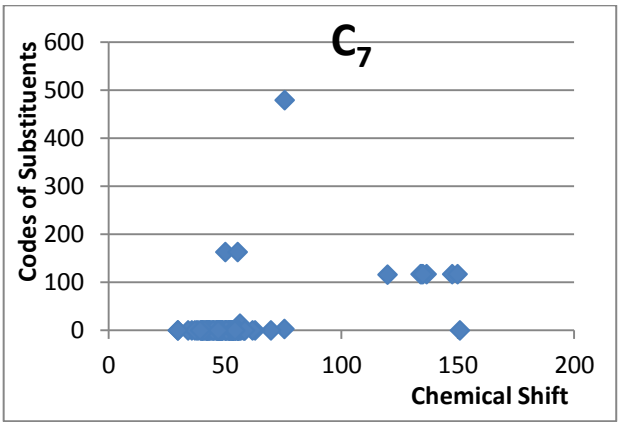


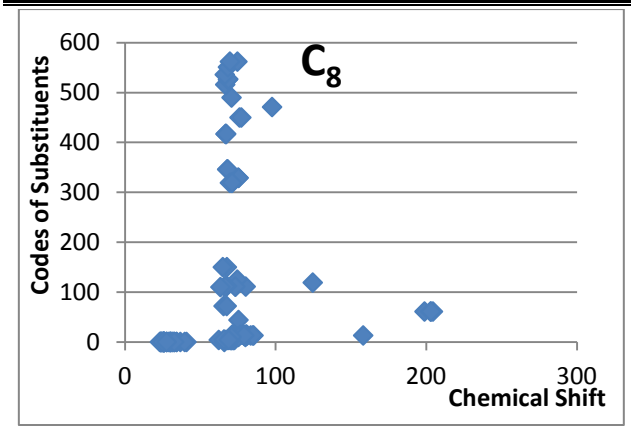


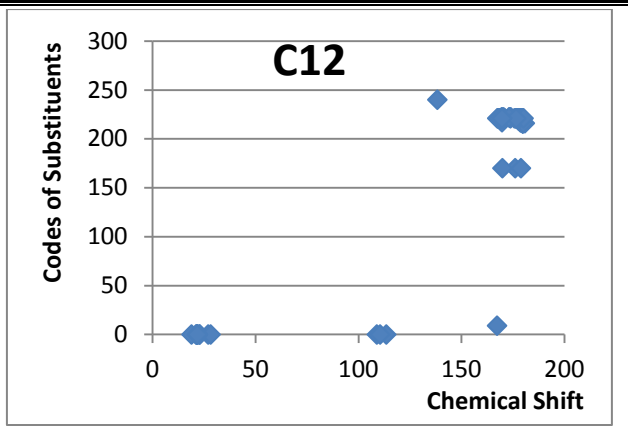


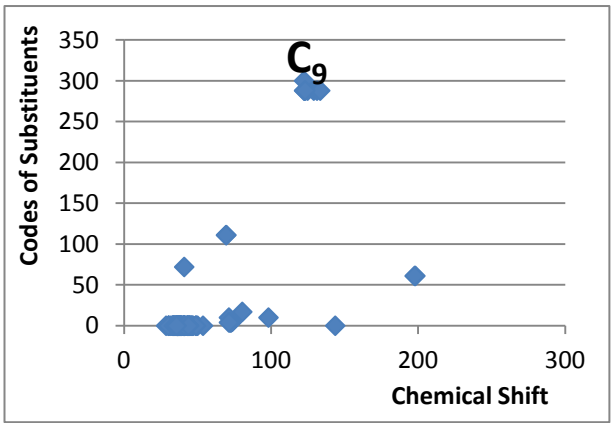


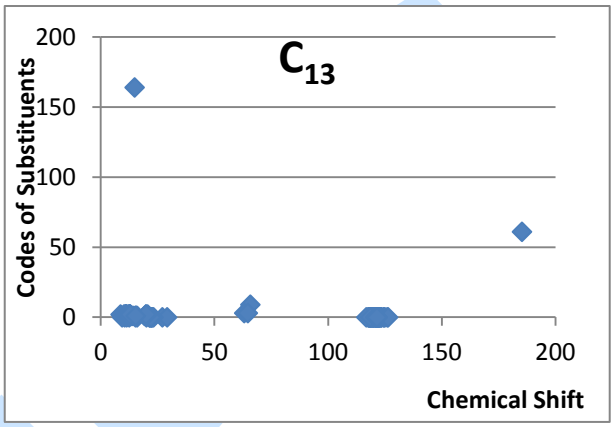


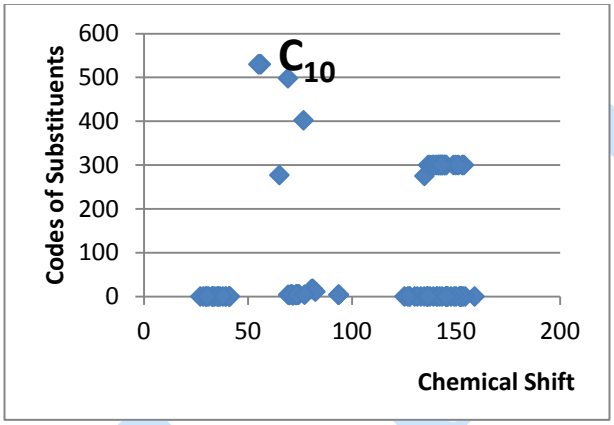


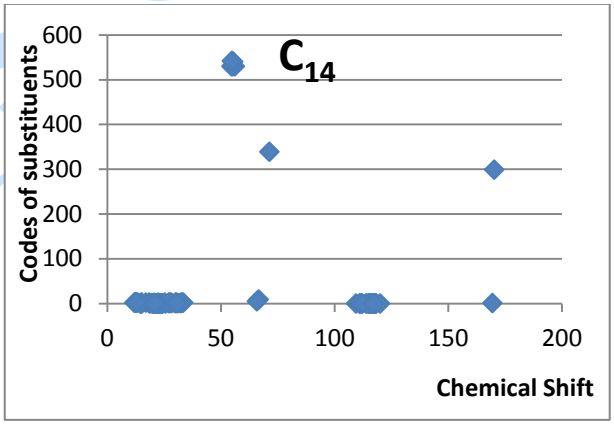

C8
Codes of Substituents
Chemical Shift
0
100
200
300
400
500
600

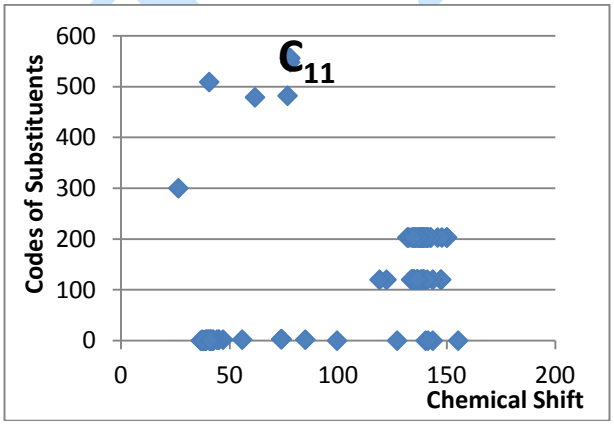

C12
Codes of Substituents
Chemical Shift
0
50
100
150
200
250
300

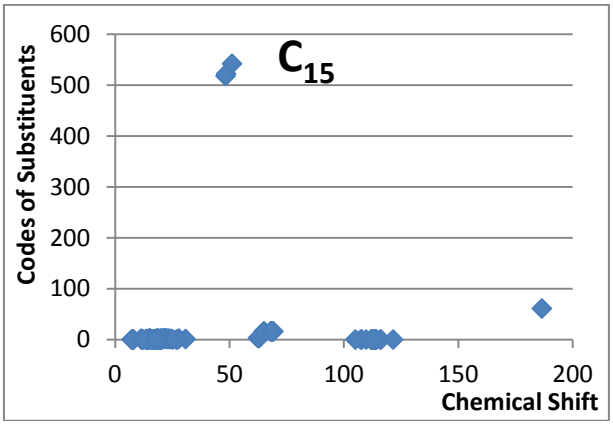

C9
Codes of Substituents
Chemical Shift
0
50
100
150
200
250
300
350


C13
Codes of Substituents
Chemical Shift
0
50
100
150
200


C10
Codes of Substituents
Chemical Shift
0
100
200
300
400
500
600


C14
Codes of substituents
Chemical Shift
0
100
200
300
400
500
600


C11
Codes of Substituents
Chemical Shift
0
100
200
300
400
500
600


C15
Codes of Substituents
Chemical Shift
0
100
200
300
400
500
600