Evans and Johnson, J Comput Sci Syst Biol 2018, 12:1
DOI: 10.4172/jcsb.1000292

Journal of
Computer Science & Systems Biology

Review Article Open Access

Machine Learning Programs Predict Saguaro Cactus Death

Evans LS* and Johnson CR

Laboratory of Plant Morphogenesis, Biological Sciences Research Laboratories, Manhattan College, Bronx, NY 10471, USA

Abstract

Objective: Determine if machine learning programs coupled with standard statistical methods can accurately predict rates of bark coverage and death of saguaro cactus plants.

Methods: Data of twelve surfaces of 1,149 saguaro cacti with four samplings over 23 years that provided more than 55,000 data points were analyzed to predict rates of bark coverage on cactus surfaces and cactus death with three machine learning programs, *Validate Model*, *WEKA* 3.8 decision trees, and *Random Forest*.

Results: Saguaro cacti (*Carnegiea gigantea*) show extensive bark coverage and cacti with extensive bark coverage die prematurely. Over the 23-year period of study, bark coverage on all surfaces was relatively constant. Decision trees are able to predict cactus death up to 96%. Three machine learning programs used similar surface coverages to make similar predictions of future bark coverage and cactus death accurately (approximately 92%), for cacti that had overall bark coverage less than 80% on south-facing surfaces. Higher prediction accuracies were obtained for cacti with were low bark percentages. While bark coverage rates and cactus death were less accurate for cacti with higher bark percentages because cacti can remain with high bark percentages with many years prior to death. Cacti with more than 80% coverage on south-facing surfaces were accurately predicted ($p<0.05$) to be alive and dead of the 23-year period with a tracking method.

Conclusions: The combined machine learning programs coupled with standard statistical procedures accurately predicted bark coverages and cactus death with greater than 95%.

Keywords: Saguaro cacti; *Carnegiea gigantea*; Bark coverage; *Validate Model*; *WEKA* 3.8 decision trees; *Random forest*; Predict cactus death

Introduction

Saguaro cacti (*Carnegiea gigantea*), a columnar plant species native to the Sonoran Desert, exhibit extensive bark coverage and experience premature death [1]. Twenty-three additional columnar cactus species show bark coverage in the Americas [2]. Bark coverage is caused by a build-up of epicuticular waxes produced by epidermal cells [3,4]. During wax accumulation, many stomata become blocked, preventing gas exchange [2-4]. Extensive bark coverage results in cactus mortality rates of 2.3% per year even though estimates of life expectancies may be several hundred years [5-7].

Controlled exposures to UV-B light produce a buildup of epicuticular waxes in the same manner that occurs in nature [8]. Averaged over the entire year, south-facing surfaces of saguaro cacti in Tucson, Arizona (32.2°N) receive four times more sunlight than north-facing surfaces [9]. Corresponding with these sunlight exposures, south-facing surfaces of saguaros show initial bark coverage whereas north-facing surfaces are the last surfaces to show bark coverage prior to cactus death (Figure 1) [2,3,10].

In previous studies, machine learning programs predicted increases in both bark coverages on cactus surfaces and death of saguaro cacti [1,6,11]. The most accurate predictions of cactus death involved bark coverage on north-facing right troughs (NR). For example, a *WEKA* C4.5 decision tree predicted cactus death with 88.7% accuracy using only trough data [6]. Two independent machine learning methods using only NR data predicted increased rates of bark coverage (RBC) and mortality (84% accuracy) [1]. Most recently, results of logistic curve analyses (with all r^2 values above 0.95), demonstrated the sequence of bark coverage on twelve surfaces around cacti. NR were the last of the twelve surfaces to have bark coverage, after an estimated 13-year delay [11].

The current study uses data from four consecutive evaluations, including three machine learning programs, to determine the following: (1) if saguaro cacti have relatively constant RBC; (2) if bark coverage on several cactus surfaces can predict bark coverage on NR; and (3) if bark coverage can predict cactus death.

The following were hypothesized: that bark coverage increases at constant rates, bark coverage of NR can be predicted with bark coverage on other surfaces, mortality rates can be predicted with bark coverage, and mortality of Class IV cacti can be predicted with bark coverage.

Methods

Field and survey conditions

Over a 23-year period, 1,149 saguaro cacti (*Carnegiea gigantea*) were studied. All cacti were within 50 field plots in Tucson Mountain Park near Tucson, AZ (32.2°N, 111.14°W) [5]. Cacti were first selected in 1993-1994. All selected cacti were taller than 4 m and assumed to be more than 80 years old [7,12]. Physical characteristics, nearby vegetation, topographical features, and GPS coordinates were used to distinguish each cactus plant. All cacti were re-evaluated in 2002, 2010, and 2017 to provide a total of four sampling periods [5].

***Corresponding author:** Evans LS, Laboratory of Plant Morphogenesis, Biological Sciences Research Laboratories, Manhattan College, Bronx, NY 10471, USA, Tel: +7188627210; Fax: +7188627814; E-mail: lance.evans@manhattan.edu

Received November 01, 2018; **Accepted** November 30, 2018; **Published** December 07, 2018

Citation: Evans LS, Johnson CR (2018) Machine Learning Programs Predict Saguaro Cactus Death. J Comput Sci Syst Biol 12: 1-8. doi:10.4172/jcsb.1000292

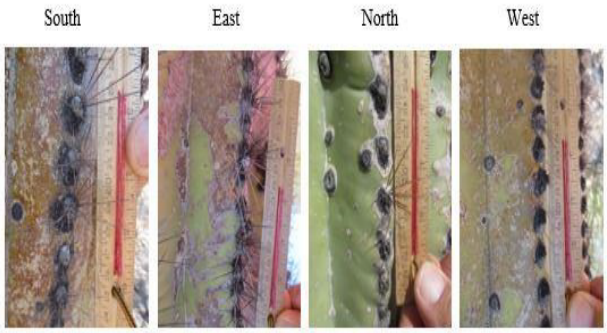

Figure 1: Visual representation of bark coverage on south, east, north, and west-facing surfaces on a saguaro cactus (*Carnegiea gigantea*).

Evaluation of cacti in field plots

Each saguaro cactus has 20 to 23 rib crests or convex protrusions [13]. Troughs are concave surfaces on the sides of each crest [14]. Usually, crests have more bark coverage percentages than troughs since crests are more exposed to sunlight than troughs. For each evaluation, the crest closest to south, east, north, and west azimuths was evaluated [3,4,15,16]. Troughs to the right and left of each crest were evaluated so that twelve surfaces were assessed for each cactus. An 8 cm vertical span, at 1.75 meters from ground level, was evaluated for percent bark coverage for each surface [3,4,15,16]. Bark coverage percentages were determined visually. Previously, similar estimates of bark coverage were obtained with visual and digital methods [6]. Percentages of bark coverage on all surfaces were entered into Microsoft Excel. The final data file titled *Master File* had 55,152 data points for the twelve surfaces of all 1,149 cacti over the four data collection periods (1993-1994, 2002, 2010, and 2017).

Classes of cactus plants

Bark coverage varies on cactus surfaces. Each cactus was placed in a class depending upon the percentage of bark coverage on south-facing crests only [5]:

Class I - Less than 20% bark coverage,

Class II - Between 21 and 49% bark coverage,

Class III - Between 50 and 80% bark coverage,

Class IV - More than 80% bark coverage,

Class V - Cactus was dead.

Estimating constant rates of bark coverage (RBC)

To determine if bark coverage occurs at constant rates, the collected data were divided into three time periods (1994-2002, 2002-2010 and 2010-2017). Rates of bark coverage (RBC) were determined to be constant when the rate of bark coverage was similar between two successive periods. Thus, the data were placed into six class changes with three time periods each. For each cactus, bark coverage percentages were converted to arcsine values to linearize the data for analysis [17]. The rate of change for each cactus was determined by:

Rate of change=X-Y/Y

where X is the arcsine bark percentage at the end of the period and Y is the arcsine bark percentage at the beginning of the period. The RBC for all samples in a class change among time periods were analyzed using a paired student's *t*-test (p=0.05).

Machine learning programs

Three machine learning programs (*WEKA* 3.8 decision trees, *Validate Model*, and *Random Forest*) were used to understand bark coverage. The *WEKA* 3.8 program generated decision trees that resulted in two distinct outcomes with a corresponding accuracy [18,19]. The *WEKA* 3.8 program implemented the ID3 algorithm to generate variables used to make predictions and produce decision trees [20]. The decision trees are visualized using the J48 algorithm.

The second machine learning program, *Validate Model*, was used to analyze RBC on cactus surfaces. *Validate Model* used a 10-fold cross-validation technique that used ten randomly selected subsets of the entire database, tested to produce a model, and validated the entire dataset to produce the final machine learning model. From *Validate Model*, rates of bark coverage (RBC) were classified as *Slow*, *Normal*, and *Fast*. Rates were considered *Slow* if the values were less than two times the standard deviation from the mean (*Normal*). Rates were considered *Fast* if the values were more than two times the standard deviation from the mean (*Normal*).

A third machine learning program, *Random Forest*, was used to confirm the results of the first two machine learning through the construction of decision trees [21].

Results

Bark coverage percentages increase at constant rates

There were six group comparisons (Class I to Class II, Class I to Class III, Class I to Class IV, Class II to Class III, Class II to Class IV, Class III to Class IV) made for twelve surfaces and three time periods (Table 1). The data show that there were no consistent differences in the RBC for any surface for the three of the time periods. Some individual differences were present for all trough surfaces. No crests showed statistically significant differences. Therefore, we conclude that bark coverage rates were relatively constant on all surfaces.

Bark coverage rates of north-facing right troughs can be predicted with bark coverage on other surfaces

To predict bark coverage on north-right troughs, three programs were utilized. The first step is to remove all dead cacti from *Master File.* The new file is titled *No Dead File*. The second step is to place the data from *No Dead File* into *Validate Model* so individual cacti could be catalogued as *Slow, Normal,* and *Fast* RBC. Three clusters were selected as predictor surfaces. Cluster 1 consisted of all surfaces except NR. Cluster 2 consisted of north-facing crests (NC) with north-facing left troughs (NL). Cluster 3 consisted of NL with west-facing left troughs (WL). After S*low, Normal,* and *Fast* groups of cacti were determined, data were processed with the *WEKA* 3.8 program. Within each cluster, comparisons were made between *Normal* and *Slow* cacti as well as between *Normal* and *Fast* cacti (Figure 2).

Comparisons of several machine learning models were used to predict bark coverage on NR with multiple predictive surfaces. The *WEKA* 3.8 and *RF* programs predicted bark coverage on NR with 89.9 to 94.9% accuracies (Table 2). Overall, analyses with *Slow* cacti produced higher accuracies than analyses with *Fast* cacti. Accuracies of the three predictive surfaces (north-facing left troughs (NL) combined with west-facing left troughs (WL), NL combined with north-facing crests (NC), and all surfaces except the NR) were similar. The numbers of *Fast* and *Slow* cacti from *WEKA* 3.8 confusion matrices were within 96% of cacti from *Validate Model.*

Changes in bark coverage (%)													
	n[a]	SC[b]	SR	SL	EC	ER	EL	NC	NR	NL	WC	WR	WL
Class[c] change													
Class I to II													
1994/2002	78	17.8	10.2*d	9.89*	12.4	5.95*	8.75*	3.88	2.22	2.22*	11.4	5.76*	3.94*
2002/2010	51	20	5.15	5.17	11	3.68	5.03	10.24	2.86	1.33	12.3	3.25	2.71
2010/2017	16	20.1	6.25	4.86	12	1.3	3.42	14.2	2.62	1.87	15.4	1.68	2.5
Class I to III													
1994/2002	26	50.2	16.7	19.6*	16.7	6.53	13.1	9.3	2.61	6.39	10	9.89*	6.12
2002/2010	28	49.3	12.2	11	21.7	4.26*	10.7*	11.3	2.18	2.82	20.9	7.5	6.0*
2010/2017	11	47.1	8.52	9.86	29.1	2.65	6.57	14.3	1.73	3.27	24.1	3.13	2.18
Class I to IV													
1994/2002	40	84.5	35.9*	36.5*	35.8	15.5*	24.1*	16.4	6.32*	5.59	30.1	18.7*	6.83*
2002/2010	22	80.5	20.75	20.1	34.9	5.09	10.8	15.6	3	2.86	31.4	5.9	3.57
2010/2017	10	83.3	36.8	30.8	38.4	3.2	17	11.3	2.6	3.1	54.2	26.3	3.6
Class II to III													
1994/2002	19	32.8	12.3*	16.7*	26.9	9.9	20.6*	5.2	3.4	4.1	9.2	12.2*	7.3*
2002/2010	34	32.7	10.9	13	24.7	7.2	10.7	13.9*	2.8	3.7	21.4	9.1	4.2
2010/2017	18	31.9	7.3	14	21.8	1.2	7.7	6.1	2.6	3	18.3	7.2	3.9
Class II to IV													
1994/2002	34	61.7	36.1	38.6*	25.6	12.4	25.5	13.9	10	8	25.3	19.6	10.5
2002/2010	45	63.2*	22.3	20.4	33.4	7	13.5	12.5	3.5	2.7	29.4	12.1*	6.1*
2010/2017	31	60.2	27.6	21.6	34.8	8.3	12.2	16.6	3.6	6.9	22.9	9.6	6.5
Class III to IV													
1994/2002	29	31.5	28.6	34.0*	23.4	7.7	21.0*	13.5	8.1*	7.2*	25.5	20.1	6.7*
2002/2010	48	31.1	26.5	24.5	33.4	12.2	18.9	25.0*	4.8	6.2	42.3	20.2	7.7
2010/2017	49	29.8	25.9	36.4	26.4	16.1	19.9	17	2.3	3.5	24	20	6

[a]n is the number of samples, [b]SC=south-facing crests, SR=south-facing right troughs, SL=south-facing left troughs, EC=east-facing crests, ER=east-facing right troughs, EL=east-facing left troughs, NC=north-facing crests, NR=north-facing right troughs, NL=north-facing left troughs, WC=west-facing crests, WR=west-facing right troughs, WL=west-facing left troughs. [c]Classes of cacti are based upon the percentage of bark coverage on south-facing crests only: Class I – has less than 20% bark coverage; Class II - between 21 and 49% bark coverage; Class III - between 50 and 80% bark coverage; Class IV - more than 80% bark coverage. [d]Values proceeded by an asterisk denote that the rate of bark coverage was statistically different between the time period and the next time period.

Table 1: Changes in bark coverage percentages for saguaro cacti (*Carnegiea gigantea*) that changed classes for 1994 to 2017.

***WEKA* Confusion Matrices**										
Time	*RF*	Predicted:	Normal	Slow	Normal	Slow	WEKA	VM	Class IV	Class IV
Interval_	Accuracy	Actual:	Normal	Slow	Slow	Normal	Accuracy	n	n	Slow %
Surfaces Used: All Surfaces Except NR[c]										
1994-2002	92.5		458	6	33	7	92.1	38	22	48.9
2002-2010	90.6		452	1	38	13	89.9	39	31	79.5
2010-2017	92.9		453	12	26	12	92.5	39	37	94.9
Surfaces Used: NL and WL										
1994-2002	92.9		468	2	36	1	92.5	37	26	70.2
2002-2010	91.7		462	5	32	8	92.1	37	33	89.2
2010-2017	89.4		471	0		0	92.7	37	34	1.9
Surfaces Used: NL and NC										
1994-2002	91.7		457	10	25	3	94.3	35	24	68.6
2002-2010	93.5		458	11	23	2	94.9	35	32	94.1
2010-2017	94.2		446	18	16	14	93.9	35	34	97.1

[a] *Validate Model*, *WEKA*, and *Random Forest* are computer algorithms that are used for analysis.
[b] *Slow* and *Normal* refers to groups of cactus plants from *Validate Model* program. *Normal* cactus plants that have average rates of bark coverage while *Slow* cacti have rates of bark coverage slower than the average rate.
[c]NC=north-facing crests, NR=north-facing right troughs, NL=north-facing left troughs, WL=west-facing left troughs. [d]Class IV cacti had more than 80% bark coverage on south-facing crests in 1994.

Table 2: *Validate Model (VM)*, *WEKA*, and *Random Forest (RF)*[a] data to predict bark coverages on north-facing right troughs for *Slow*[b] outliers for saguaro cacti (*Carnegiea gigantea*) alive from 1994 to 2017.

Class IV cacti comprised 48 to 54%, 75 to 76.8%, and 88.6 to 93.6% for 1994-2002, 2002-2010, and 2010-2017 respectively, of all *Fast* cacti from *WEKA* 3.8 decision trees. In a similar manner, Class IV cacti comprised 48 to 72%, 79 to 94%, and 91 to 97% for 1994-2002, 2002-2010, and 2010-2017 respectively, of all *Slow* cacti from *WEKA* 3.8 decision trees. Overall, most outliers in these analyses were Class IV cacti.

Mortality rates can be predicted with bark coverage data

The *Master File* was used to predict cactus death. For the periods (1994-2002, 2002- 2010, 2010-2017), each cactus was placed in its class accordingly. Considering each class individually, *WEKA* 3.8 decision trees were used to determine whether a cactus was alive or dead by the end of the period.

The *WEKA* 3.8 program was used to predict cactus death with data of several cactus classes. *WEKA* 3.8 decision trees predicted death of Class I cacti with accuracies of 89.2%, 83.5%, and 97.1% from 1994-2002, 2002-2010, and 2010-2017, respectively (Table 3; Figure 3).

Decision trees predicted death of Class II cacti with accuracies between 64.9 and 95.2% (Figure 4). Decision trees predicted death of Class III cacti with accuracies between 57.1 and 92.5%. Decision trees predicted death of Class IV cacti with relatively low accuracies between 68.9 and 77.4%.

Mortality of Class IV cacti can be predicted with bark coverage data

Within the *Master File*, 1994's Class IV cacti were removed and saved. The next step was to remove data of cacti that died in 2002 and 2010. The resultant file was called *Class IV Selection*. The last step was to separate data from *Class IV Selection* into cacti that remained alive in 2017, and those that died by 2017 (after 2010). Cacti living in 2017 were labelled Class IV Alive, and cacti that died by 2017 were labelled Class IV Dead. A series of *t*-test comparisons were made between the two files for bark coverage percentages on all surfaces for 1994, 2002, and 2010. Class IV Dead had higher bark coverage than Class IV Alive for all surfaces, except for (1) south crests in 2002 and 2010; (2) NR in 1994; (3) west-facing crests for 2010; and (4) WL for 1994 and 2002 (Table 4). Although *Validate Model* and the *WEKA* 3.8 program were not accurate for predicting death of Class IV cacti accurately, death of Class IV cacti can be predicted based upon bark coverages from 1994 through 2017.

Discussion

Saguaro cacti can live several hundred years [7,12]. Bark coverage on saguaro cacti did not occur, or was rare, prior to the 1950's [10]. Thus, bark coverage on saguaros is a relatively recent phenomenon. The purpose of this study was to closely examine bark coverage and cactus death within a population of saguaro cacti. Turner showed that 158, of a total of 208, cacti were taller than 4.2 m in 1962. In the same plot of 1988, only 27 of 168 cacti were taller than 4.2 m [22]. By subtraction, a minimum of 131 saguaros, taller than 4.2 m, died over the 27-year period. In addition, Turner and Funicelli demonstrated that 16% of the saguaro cacti in their study plots died between 1990-2000 [23]. O'Brien et al. demonstrated that the oldest of 20,372 saguaros in Saguaro National Park was less than 110 years old [24]. The above data and the data of the current study are coincident with high mortality rates [5]. These increases in morbidity and mortality are inconsistent with the lifespan noted above [7].

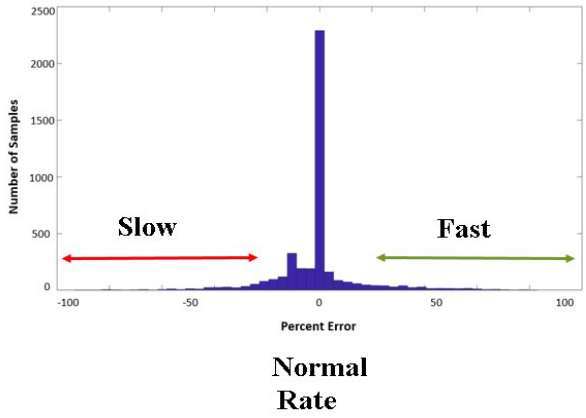

Figure 2: Image of the print out of data from *Validate Model*. The print out shows the distribution of rates of bark coverage (RBC) on a surface. *Normal* rates are defined as rates within two standard deviations of 0% error. Cacti with *Fast* rates are values that are more than two standard deviation units above 0. Cacti with *Slow* rates are values that are less than two standard deviation rates below 0.

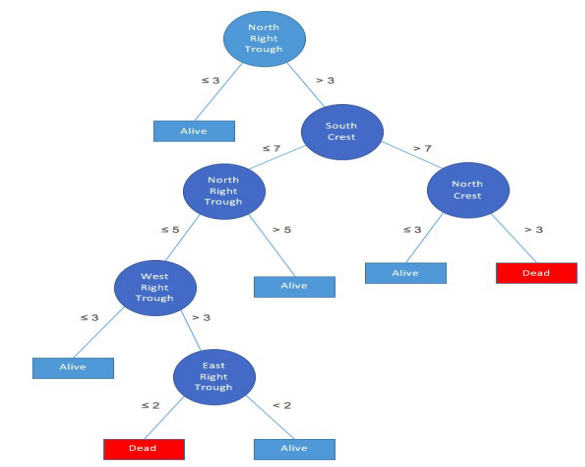

Figure 3: *WEKA* decision tree predicting death among Class I cacti from 1994-2002.

As stated previously, saguaros have extensive bark coverage prior to experiencing premature death [1]. Bark coverage on south-facing surfaces first occurs on south-facing surfaces while north-facing surfaces are the last surfaces to have bark coverage before cactus death [2,3,10]. For example, south crests of Class I cacti had an 80% increase in bark coverage, while east and west crests increased between 35 and 50%. Concurrently, NC increased 11 to 15%, while NR and NL increased only 2.6 and 3.1% respectively. Similar large differences occurred for Class II cacti that moved to Class IV. South crests increased 60%, east and west crests changed between 25 and 35%, while north crests changed 12 and 16%, and NR and NL changed only 3.6 and 6.9% respectively. The current data are incomplete in accordance with previous data from 1994, 2002, and 2010 that demonstrate the delay in bark coverage from south-facing surfaces to north-facing surfaces [11].

Saguaro cacti younger than 80 years old have little bark coverage. Bark coverage begins slowly on a surface but eventually increases rapidly. Surfaces above 90% bark coverage increase slowly thereafter.

Class[a] predicted	Predicted Accuracies (%)		Confusion matrices		Accuracies (%)
			Predicted alive	Predicted dead	
I 1994 – 2002	n=231 NR[b] 3, SC 7, NC 3, WR 3, ER 2	actual alive	203	22	89.2
		actual dead	3	3	
I 2002-2010	n=231 NR[b] 3, SC 7, NC 3, WR 3, ER 2	actual alive	71	11	83.5
		actual dead	0	3	
I 2010-2017	n=35 no tree	actual alive	34	1	97.1
		actual dead	0	0	
II 1994-2002	n=111 NC 35, NC 95, EL 10	actual alive	94	12	84.7
		actual dead	0	5	
II 2002-2010	n=37 SC 35	actual alive	23	9	64.9
		actual dead	0	4	
II 2010-2017	n=84 no tree	actual alive	80	4	95.2
		actual dead	0	0	
III 1994-2002	n=81 no tree	actual alive	70	11	86.4
		actual dead	0	0	
III 2002-2010	n=14 SR 10, EC 40	actual alive	8	3	57.1
		actual dead	0	3	
III 2010-2017	n=107 NR 22, SL 40	actual alive	99	4	94.5
		actual dead	0	4	
IV 1994-2002	n=463 WL 75, NR 55, EL 95, SL 99, NL 35	actual alive	284	113	68.9
		actual dead	35	31	
IV 2002-2010	n=447 NL 50, SR 80, WC 96, ER 95, NC 55, NR 30	actual alive	272	67	77.4
		actual dead	74	34	
IV 2010-2017	n=425 SL 95, NL 45, NC 85, WL 45, NC 65	actual alive	279	70	72.6
		actual dead	37	49	

[a] Cactus plants were placed into bark coverage classes based upon the percentage of bark coverage on south-facing crests only. Class I - less than 20% bark coverage; Class II - between 21 and 49% bark coverage; Class III - between 50 and 80% bark coverage; Class IV - more than 80% bark coverage.
[b] Cactus surfaces: SC=south-facing crests, SR=south-facing right troughs, SL=south-facing left troughs, EC=east-facing crests, ER=east-facing right troughs, EL=east-facing left troughs, NC=north-facing crests, NR=north-facing right troughs, NL=north-facing left troughs, WC=west-facing crests, WR=west-facing right troughs, WL=west-facing left troughs.
[c] SC 7 refers to the limits for the decision trees. For example, SC 7, the decision divided cactus plants based 7% on south crests.

Table 3: Using WEKA to predict death of saguaro cacti (*Carnegiea gigantea*) with various predictive surfaces over each time interval. For simplicity, only the first 5 steps of the decision trees are shown for the predictor surfaces.

2017													
Time	Status	SC[b]	SR	SL	EC	ER	EL	NC	NR	NL	WC	WR	WL
1994	Alive	95.8**	47.2**	49.3**	71.7*	30.7*	50.1**	41.4**	10.4	18.4	57.2**	34.7*	14.4
	Dead	98.4**	67.5**	65.5**	82.0*	42.0*	68.2**	57.0**	15.4	26.2	70.0**	47.2*	16.8
2002	Alive	98.3	68.5**	70.7**	82.0**	42.0**	67.2**	55.6**	19.1*	27.1**	74.9*	55.9*	30.1
	Dead	98.6	83.9**	86.2**	91.8**	68.7**	83.4**	73.9**	27.4*	41.4**	85.2*	66.8*	37.2
2010	Alive	99.1	81.4**	85.4*	89.4**	62.7**	79.4**	75.0**	33.3**	38.5**	92	71.8**	54.1*
	Dead	98.6	92.9**	92.7*	96.9**	78.6**	90.9**	91.2**	49.7**	56.7**	94.4	84.7**	63.8*

[a] Classes of cacti are based upon the percentage of bark coverage on south-facing crests only: Class IV - more than 80% bark coverage.
[b] SC=south-facing crests, SR=south-facing right troughs, SL=south-facing left troughs, EC=east-facing crests, ER=east-facing right troughs, EL=east-facing left troughs, NC=north-facing crests, NR=north-facing right troughs, NL=north-facing left troughs, WC=west-facing crests, WR=west-facing right troughs, WL=west-facing left troughs.

Table 4: Mean data of saguaro cacti (*Carnegiea gigantea*) that began as Class IV[a] in 1994 and either remained alive into 2017 or reached mortality in 2017. A t-Test was performed on raw data values of cacti that remained alive in 2017 and cacti that died in 2017 for each evaluation period. Black values proceeded by no asterisk are statistically similar. Values proceeded by one asterisk were significantly different by a t-Test value between .01 and .05. Values proceeded by two asterisks were significantly different by a t-Test value of less than .01.

So, the increase in bark coverage reflects a logistics curve [11]. Data of Class I cacti that moved to Classes II, III, and IV may reflect a logistics curve (Figure 5). The low slope for south crests may be projected to the lower portion of a logistics curve. The higher slopes of east and west crests may be projected to the initial increase in the logistics curve. The south crest's slope may be projected to the upper part of the logistics curve.

All three machine learning models provided very similar results. The *WEKA* 3.8 program and *Random Forest* accuracies from all samples were within 4.2% (Table 2). Moreover, data of *Fast* and *Slow* cactus groups from *Validate Model* were within 2.9% of *WEKA* 3.8 decision tree values. The greatest error between the results of the machine learning models occur because of outliers. The majority (88%) of outliers were Class IV cacti. Therefore, most of the *Normal* cacti were from Classes I, II, and III.

Over the past several decades, many new and innovative decision tree programs have been developed. Decision trees have been used for a large number of purposes in making predictions. For example, field guides that are used to assist in the identification of wild bird and plant species serve as a form of decision tree [20,25]. Machine learning decision trees enable conclusions with large data sets in a wide range of topics such as board games, astronomy, petrochemistry, cancer research, phylogenetics and among many others [26-30].

Decision trees that are used to address large databases have no appeal to a conceptual model. Therefore, decision trees may serve to validate models and concepts held by researchers, based upon current knowledge. For example, a previous *WEKA* 3.8 decision tree analysis of saguaro cactus data, from the 1994 and 2002 sampling periods, selected north-facing surfaces as the first indicator surfaces [6]. Data from the 2010 survey revealed that if NL had more than 70% bark coverage, and NR had more than 65% bark coverage, *WEKA* 3.8 decision trees predicted with 85% accuracy that a cactus would die within eight years.

The above predictions, from three time periods (1994, 2002, and 2010), are supported by the data herein that include a survey from 2017. Summation data for all three north-facing surfaces of Class IV Alive cacti were 71 to 74% of Class IV Dead for 1994, 2002, and 2010. In addition, summation data of east troughs, north troughs, and west troughs of Class IV Alive cacti were 72 to 79% of Class IV Dead for the same three time periods (Table 5). The 70 to 79% lower percentages on troughs may represent the lag time of bark coverage on troughs compared to crests. In addition, a sum of all twelve surfaces was created for Class IV Alive and Class IV Dead. Class IV Dead cacti had a sum of 656 in 1994 (23 years prior to 2017), and Class IV Alive cacti had a sum of 691 in 2002. Do these data suggest that Class IV Alive cacti will die within 23 years prior to 2025?

Many *WEKA* 3.8 decision trees were generated for this research; however, machine learning programs have some limitations. Although *Random Forest* produced confusion matrices and had high accuracies, the program does not produce decision trees. Machine learning models used by *WEKA* 3.8 generate decision trees using a ten-fold, cross-validation technique. This technique uses 90% of the dataset for training and utilizes the remaining 10% of cacti to generate trees, confusion matrices, and accuracies. If a dataset has few cacti, the *WEKA* 3.8 program may not have enough information to provide high accuracies. For example, Class I cacti from 20102017 had only 35 cacti available for examination. Only one cactus died, while 34 remained alive. For this case, the proportion of alive to dead cacti was inadequate to train the *WEKA* 3.8 program to produce an accurate decision tree.

Predicting death of Class IV cacti from 1994-2017 using the *WEKA* 3.8 procedures may prove to be more successful than doing so from 1994-2002, 2002-2010, and 2010-2017. Attempts to predict the fate of Class IV cacti from 1994-2017 using the *WEKA* 3.8 program gave low accuracies. The low accuracies most likely resulted from the slow increases in bark coverage exhibited by already unhealthy Class IV cacti. As shown in this study, the tracking of bark coverage from period to period (1994-2002, 2002-2010, and 2010-2017), provided an accurate way to differentiate between alive and dead cacti in 2017. Differentiation between alive and dead cacti of Class IV began in 1994. To our knowledge, this is the first publication that predicts cactus health with three independent machine learning models. Therefore, machine learning models provide powerful tools to address complex issues with large data bases.

Author Contributions

LSE was involved in the initial experimental designs, field work in evaluating the cacti, helping to format the excel databases, evaluation of

Year	Bark Coverage all surfaces	Coverage differences alive/dead (%)	Bark Coverage north surfaces only (%)	Coverage differences alive/dead (%)	Bark Coverage trough east north west	Coverage differences alive/dead (%)
Class IV Alive						
1994	521	79.4	70.2	71.2	165	73.3
2002	691	81.9	102	71.3	241	72.6
2010	862	87.1	147	74.2	343	79.6
Class IV Dead						
1994	656		98.6		226	
2002	845		143		332	
2010	991		198		431	

[a] Classes of cacti are based upon the percentage of bark coverage on south-facing crests only: Class IV - more than 80% bark coverage.

Table 5: Summations and percentages of bark coverages for *Class IV*[a] *Alive* and *Class IV Dead* cacti of saguaro cacti (*Carnegiea gigantea*).

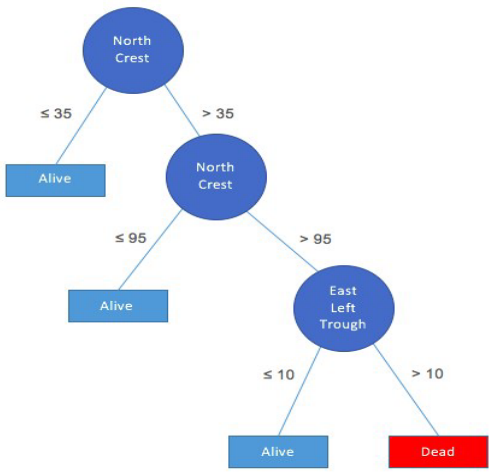

Figure 4: *WEKA* decision tree predicting death among Class II cacti from 1994-2002.

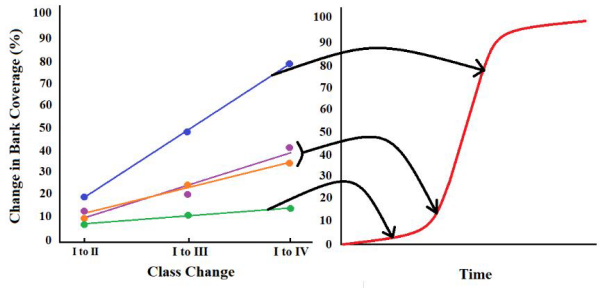

Figure 5: Relationships between changes in bark coverages for crests of Class I cacti that changed to Classes II, III, and IV. All slopes had r^2 values above 0.96 and the slopes were 30.1, 14.5, 11.5, and 3.5 for south crests (green circles), west crests (orange circles), east crests (purple circles), and north crests (blue circles), respectively. Note the black solid arrows that indicate the possible location of these data with a logistics curve.

the data, writing of the manuscript. CRJ was involved in the field work to evaluate cacti, entered all data into excel data sheets and proofread them, entered the data into the various machine learning programs, and did the initial evaluations of the results of the machine learning programs, assisted with the editing process and creation of graphics. The authors are the sole originators of the data and have no conflicts of interest. The funding for this research was from the Catherine and Robert Fenton Endowed Chair to LSE and is the only funding source. To be clear, we received no funding or any other type of support from any commercial sources and we have no conflicts of interest with this research and/or this report. This manuscript has not been submitted elsewhere in any form.

Acknowledgements

The authors are grateful for support by the Catherine and Robert Fenton Endowed Chair to Lance S. Evans. The authors thank Mr. Mark Brosseau and other personnel at Tucson Mountain Park (TMP), Pima County Parks, and Recreation Department for a Special Use Permit for many years to revaluate the populations of saguaro cacti in Tucson Mountain Park. The authors thank Ms. Lilliana McHale and Dr. Heidi Laudien for the construction and editing of this manuscript.

References

1. Evans LS, De Bonis M (2015) Predicting morbidity and mortality of Saguaro cacti (*Carnegiea gigantea*). The Journal of the Torrey Botanical Society 142: 231-239.
2. Evans LS, Macri A (2008) Stem surface injuries of several species of columnar cacti of Ecuador. The Journal of the Torrey Botanical Society 135: 475-482.
3. Evans LS, Cantarella VA, Stolte KW, Thompson KH (1994) Phenological changes associated with epidermal browning of saguaro cacti at Saguaro National Monument. Environmental and Experimental Botany 34: 9-17.
4. Evans LS, Cantarella VA, Kaszczak L, Krempasky SM, Thompson KH (1994) Epidermal browning of saguaro cacti (*Carnegiea gigantea*). Physiological effects, rates of browning and relation to sun/shade conditions. Environmental and Experimental Botany 34: 107-115.
5. Evans LS, Young AJ, Harnett SJ (2005) Changes in scale and bark stem surface injuries and mortality rates of saguaro (*Carnegiea gigantea*) cacti population in Tucson Mountain Park. Canadian Journal of Botany 83: 311-319.
6. Evans LS, Boothe P, Baez A (2013) Predicting morbidity and mortality for a saguaro cactus (*Carnegiea gigantea*) population. The Journal of the Torrey Botanical Society 140: 247-255.
7. Steenbergh WF, Lowe CH (1997) Ecology of the Saguaro II reproduction, germination, establishment, growth, and survival of the young plant. National Park Service Monograph Series Eight.
8. Evans LS, Sullivan J, Lim M (2001) Initial effects of UV-B radiation on stem surfaces of *Stenocereus thurberi* (organ pipe cacti). Environmental and Experimental Botany 46: 181-187.
9. Geller G, Nobel P (1984) Cactus ribs: Influence of PAR Inception and CO_2 uptake. Photosynthetica 18: 482-494.
10. Evans LS, Howard KA, Stolze E (1992) Epidermal browning of Saguaro Cacti (*Carnegiea gigantea*): Is it new or related to direction?. Environmental and Experimental Botany 32: 357-363.
11. De Bonis M, Barton L, Evans LS (2017) Rates of bark coverage of surfaces of saguaro cacti (*Carnegiea gigantea*). The Journal of the Torrey Botanical Society 131: 236-248.
12. Pierson E, Turner R (1998) An 85-year study of saguaro (*Carnegiea gigantea*) demography. Ecology 79: 2676-2693.
13. Anderson EF (2001) The cactus family. Timber Press, Portland, p: 776.
14. Gibson A, Nobel P (1986) The cactus primer. Harvard University Press, Cambridge, MA, USA, p: 286.
15. Evans LS, Sahi V, Ghersini S (1995) Epidermal browning of saguaro cacti (*Carnegiea gigantea*): Relative health and rates of surficial injuries of a population. Environmental and Experimental Botany 35: 554-562.
16. Evans LM, Zugermayr M, Young AJ (2003) Changes in surface and mortality rates of saguaro (*Carnegiea gigantea*) cacti over a twelve-year paper. The Journal of the Torrey Botanical Society 130: 238-243.
17. Snedecor G, Cochran W (1967) Statistical methods. 6th edn. The Iowa State University Press, Ames.
18. Hall M, Frank E, Holmes G, Pfrahringer B, Reutemann P, et al. (2009) The WEKA data mining software: An update. SIGKEE Explorations 11: 10-18.
19. Hyafil L, Rivest R (1976) Constituting optimal binary decision trees in NP-complete. Information Processing Letters 5: 15-17.
20. Hitchcock A (1971) Manual of the grasses of the United States. Dover Publishers, New York, USA, p: 1051.
21. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE) arguments against avoiding RMSE in the literature. Geoscientific Model Development, pp: 247-1250.
22. Turner RM (1992) Long term saguaro population studies at Saguaro National Monument. In: Stone CP, Bellantoni ES (eds.). Proceedings of the symposium on research in Saguaro National Monument. Southwest Parks and Monuments Association, Tucson, Arizona.
23. Turner D, Funicelli C (2000) Ten-year resurvey of epidermal browning and population structure of saguaro cactus (*Carnegiea gigantea*) in Saguaro National Park. United States Geological Survey Technical Report, Tucson, Arizona, p: 30.
24. O'Brien K, Swann D, Springer A (2011) Results of the 2010 saguaro consensus at Saguaro National Park. National Park Service, US Department of Interior, Tucson, Arizona, p: 49.

25. White A, Dyer RA, Sloane BL (1941) The succulent Euphorbiace (Southern Africa). Abbey Garden Press, Pasadena, CA, USA, p: 937.

26. Samuel AL (1959) Some studies in machine learning using the game of checkers. IBM Journal of Research and Development 3: 210-229.

27. Salzberg S, Chandar R, Ford H, Murthy S, White R (1995) Decision trees for automated identification of cosmic-ray hits in Hubble Space Telescope images. Publications of the Astronomical Society of the Pacific 107: 279-288.

28. Blurock ES (1995) Automated learning of chemical concepts: Research octane number and molecular substructures. Journal of Computational Chemistry 19: 91-99.

29. Decaestecker C, Remmelink M, Salmon J, Camry I, Goldschmidt D, et al. (1995) Methodological aspects of using decision trees to characterize leiomyomatous tumors. Cytometry 24: 83-92.

30. Brammer G, Williams TL (2010) Using decision trees to study the convergence of phylogenetic analyses. IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, p: 8.