International Journal of Computer (IJC)
ISSN 2307-4523 (Print & Online)

http://ijcjournal.org/

Using Decision Tree to Predict Armed Conflicts in Sudan

Osman Abbas*[a], Dr. Mohamed Elhafiz Mustafa Musa[b], Dr. Siddig Balal Ibrahim[c]

[a]*MIC Corporation ,Khartoum, Sudan, Postcode 11111*
[b]*College of Computer Science and Information Technology, Sudan University of Science and Technology, Sudan.*
[c]*College of Business Studies, Sudan University of Science and Technology, Sudan .*
[a]*Email:Osman.sd@live.com*

Abstract

Security is a state where values, beliefs, democratic way of life, institutions of governance, welfare and well-being as a nation and people are permanently protected. There are many ways to predict threats which can affect this state of security [1]. The present study aimed at finding a way to predict armed conflicts in Sudan using decision trees. The main problem in this paper is that the armed confrontations are difficult to predict, because there are many elements interfere in deciding whether the conflict will be triggered or not. So this paper solved this problem using Decision tree.

Keywords: Threats; Armed conflict; Prediction and Decision trees

1. Introduction

This study investigates the effectiveness of decision trees in predicting armed conflicts, for the objectives of this research this paper discusses field operations that took place in different fronts in Sudan in the military operations areas, and considers it as a case study. The paper using decision tree because it has the advantage of being easy to interpret, which makes it practical and useful and it can be useful for generating practical solutions to complicated problems. Decision trees [2] make no assumptions about the data which means that the induction algorithms do not rely on any other information other than that which exists in the data. They can handle both categorical and continuous variables. The purpose of this study is to determine the effect of decision trees in predicting security threats in the armed conflict between Sudan government and rebels before and after separation.

---

* Corresponding author.
E-mail address: Osman.sd@live.com

There are many previous works in using decision trees in prediction whether it is on Arabic or international level , but the contribution which this paper has done is that it is the first time using decision trees in predicting armed conflicts in Sudan, so the military institution can benefit from the result and evaluate it even more. This paper will contain addition to this introduction**:** Related works, Predicting armed conflicts, Decision Trees, Methods (The main idea, Collecting Data and Analyzing data), Results and discussion, Conclusion and future work and References.

2. Related Works

one of the most related researches was the research which took place in Imperial College, University of London, for master of military science (Using decision trees to direct the planning thought-process: an enhancement to the planning methodology). The motivation for the thesis is that battlefield uncertainty, especially that posed by the enemy, is inadequately considered in the planning process. Several observations led to this perception. Planning staffs appear to have unconsciously built plans based on enemy intentions rather than enemy capability [1]. Planners tend to identify the "Most Likely Enemy Course of Action" fairly early in the planning process, usually at a stage where intelligence has not developed the situation adequately. Although this is only a subjective judgment to facilitate planning, planners often take it as fact and base friendly courses of action solely on what is perceived to be the enemy intention. This tendency to exclude the possibility of other enemy intentions locks planners into a single train of thought. Given the uncertainty, it is quite remarkable that most friendly courses of action (COAs) do not consist of more contingency or branch plans. Soldiers often lament that the plan does not survive the first contact with the enemy. This being the perception, perhaps we should make plans that are more robust against enemy reactions. Another related research was conducted in Claremont Graduate University for using Decision Trees to Predict Crime Reporting.

The notion is that Crime reports are used to find criminals, prevent further violations, identify problems causing crimes and allocate government resources. Unfortunately, many crimes go unreported. The National Crime Victimization Survey (NCVS) comprises data about incidents, victims, suspects and if the incident was reported or not [2]. The research using the NCVS is limited to statistical techniques resulting in a limited ‘view’ of the data. The goal is to use decision trees to predict when crime is reported or not. It compares decision trees that are built based on domain knowledge with those created with three variable selection methods. The research concludes that using decision trees leads to the discovery of several new variables to research further.

Another related research was (Decision Tree Algorithms Predict the Diagnosis and Outcome of Dengue Fever in the Early Phase of Illness [3]. This study shows a proof-of-concept that decision algorithms using simple clinical and haematological parameters can predict diagnosis and prognosis of dengue disease, a finding that could prove useful in disease management and surveillance.

Another research was (Predicting cesarean delivery with decision tree models). The purpose of this study was to determine whether decision tree- based methods can be used to predict cesarean delivery [4]. Decision tree receiver operating characteristic curve areas were as follows: nulliparous, 0.82; parous, 0.93. Logistic receiver operating characteristic curve areas were as follows: nulliparous, 0.86; parous, 0.93. Decision tree methods and

logistic regression methods used similar predictive variables; however, logistic methods required more variables and yielded less intelligible models. Among the 6 decision tree building methods tested, the strict minimum message length criterion yielded decision trees that were small yet accurate. Risk factor variables were identified in 676 nulliparous cesarean deliveries (69%) and 419 parous cesarean deliveries (47.6%).

3. Predicting armed conflicts

The intelligence community (IC) is asked to predict outcomes [5] that may often be inherently unpredictable and is blamed for the inevitable forecasting failures, be they false positives or false negatives. Intelligence agencies are under intense pressure to predict the arguably unpredictable. The core consumers of intelligence analysis, policy makers in the executive and legislative branches, want the intelligence community (IC) to tell them something they didn't already know. Research [6] on experts in unclassified settings suggests that the outcomes that governments ask intelligence analysts to forecast range from the very difficult to predict to the virtually impossible to predict in periods of stability. Experts are hard pressed to out-predict simple extrapolation algorithms, and in periods of turbulence, experts are hard pressed to predict random guessing strategies. Here in Sudan armed conflicts are one of the most annoying threats that authorities are trying to predict so as to take suitable arrangements to confront it. Military officers recognize that one of the key activities in battle is the management of uncertainty and that the key uncertainty on the battlefield is generated by the enemy. War is a contest of wills [1]. To bend the enemy to our will, we must fashion our actions to frustrate those of the enemy. It follows that a thorough grasp of enemy intentions is essential. The adversary understands this too. They will protect their intentions from discovery for as long as possible.

Indeed, they will even take steps to actively [7] deny them to us by the cunning use of deception. Try as we may, we cannot always discover enemy intentions regarding to what mentioned above there are many researcher conducted for this purpose. Military intelligent is the main agency which responsible for information about military operations and armed enemies if they are rebels fighting inside Sudan borders or they are enemies from abroad .So this paper is trying to give the intelligent community a predicting tool which can predict security threats in operations' areas to decide whether there will be threat of armed confrontation or not.

4. Decision Tree

A decision tree [8]; depicts rules for dividing data into groups. The first rule splits the entire data set into some number of pieces, and then another rule may be applied to a piece, different rules to different pieces, forming a second generation of pieces. In general, a piece may be either split or left alone to form a final group. The tree depicts the first split into pieces as branches emanating from a root and subsequent splits as branches emanating from nodes on older branches. The leaves of the tree are the final groups, the un-split nodes. For some [1] perverse reason, trees are always drawn upside down, like an organizational chart. For a tree to be useful, the data in a leaf must be similar with respect to some target measure, so that the tree represents the segregation of a mixture of data into purified groups. Prediction is often the main goal of data analysis. Creating a predictive model is not as automatic as one might hope. Data often arrive at the analyst's door with lots of variables. The baggage sometimes includes a dictionary that makes uninteresting reading. Yet the analyst must find something interesting in the data. Most of the variables are redundant or irrelevant and just get in the way.

A preliminary [9]. task is to determine which variables are likely to be predictive. This paper is trying to use this predicting tool to forecast the probability of armed confrontation between Sudan government and other armed groups, so the intelligent community can do their final assessment according to decision tree's outputs.

5. Method

The primary methods for achieving the objectives of the research and as illustrated in (Figure 1) the proposed architecture are designing the main frame for the decision tree which can predict whether the situation in the ground will lead to armed conflict or not. Then collecting data from resources about the previous incidents at the military operations fields whether it led to armed conflict or not. Then using SPSS program to analyze the collected data and design the decision tree which can predict the future threats.

5.1 The main idea

Any armed conflict has some elements that causing it, or at least act as stimulation factors, so the paper identified 6 main elements (variables) which are: Land, Weapon, Season, Time, Force and Tension. and then identified the elements which are derived from the main elements as follows : Land : desert, forest and mountains , Weapon: good, average and bad, Season : summer, winter and fall , Time : day and night , Force : battalion, company and platoon . Tension: yes and no. As illustrated in figure (1), and then we put it in decision tree chart and used SPSS v.21 to get results after entering hundreds of records.

5.2 Collecting Data

Sudanese armed forces' HQ is the best place where can found accurate data about previous conflicts which took place in different fronts in operation fields in Sudan. So after getting all data that needed about different conflicts in different parts in Sudan. It has been found that the best way to get benefit from this data is to analyze it according to the elements which identified earlier, as illustrated in figure (2).

5.3 Analyzing data

The algorism which we've been used is ID3 algorithm. The core idea of ID3 algorithm is information entropy which selects the property of the smallest information entropy as the category property, recursively expand branch of the tree to construct the decision tree.

The basic principles of ID3 algorithm as follows: assuming $E = F1 \times F2 \times \ldots \times Fn$ is a finite n-dimensional vector space, $Fi(i = 1,2, ..., n)$is a finite set of discrete symbols, the elements $e = (V1, V2, ..., Vn)$ in E is called the sample, in which $Vi \in Fi$, $i = 1,2, ..., n$. Supposing the number of collection of positive examples and negative examples in the collection of vector space E is the size of p and n, ID3 algorithm is based on the following two assumptions:

1) A correct decision tree in the vector space E, the probability for any classification of the sample with the E line in the probability of positive and negative cases. See Figure (3).

2) The amount of information required for a decision tree can make the right type of judgment to determine. [10].

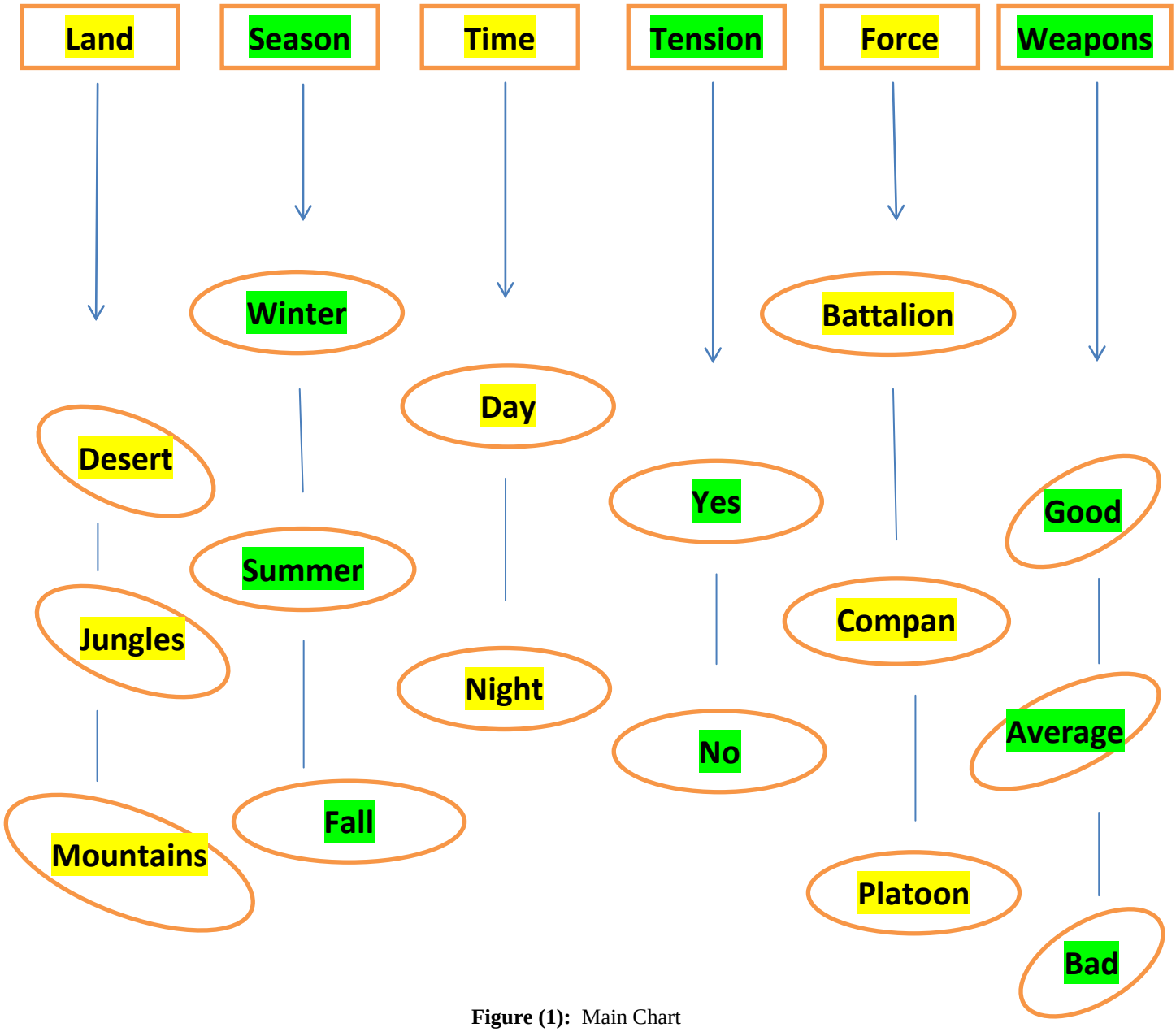
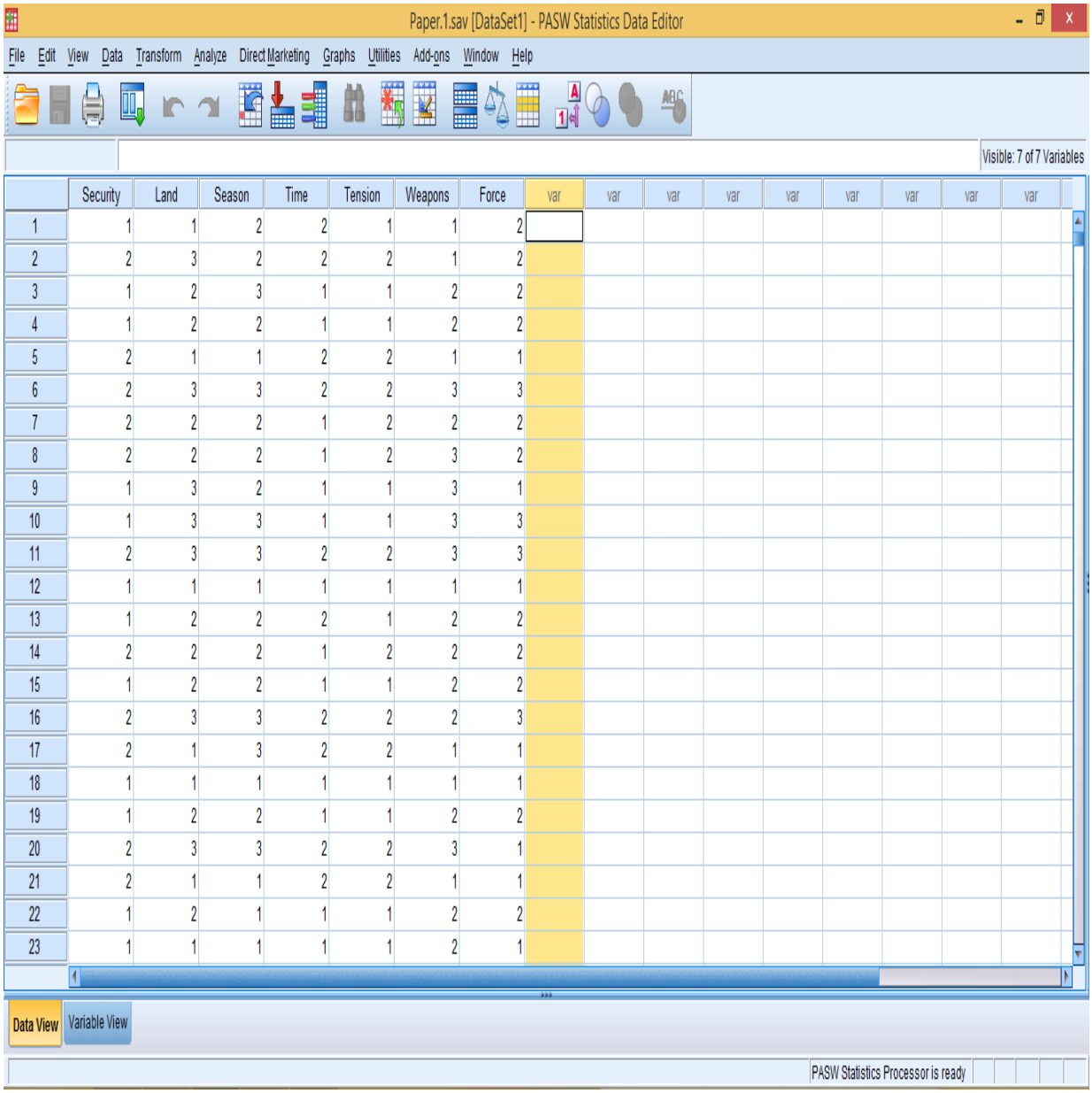


Figure (1): Main Chart

6. Results and Discussion

It has been found that the most effective variables in causing security threats are: tension, season, force and weapon. But there are exceptions according to the importance of an element comparing with other elements in every case , as illustrated in figure (4), so the research has found that the best way to get accurate results from this method is to have an accurate data . In some cases it will be difficult to find all elements, but this will not stop the program from getting the results, though it may affect the percentage of accuracy.



	Security	Land	Season	Time	Tension	Weapons	Force
1	1	1	2	2	1	1	2
2	2	3	2	2	2	1	2
3	1	2	3	1	1	2	2
4	1	2	2	1	1	2	2
5	2	1	1	2	2	1	1
6	2	3	3	2	2	3	3
7	2	2	2	1	2	2	2
8	2	2	2	1	2	3	2
9	1	3	2	1	1	3	1
10	1	3	3	1	1	3	3
11	2	3	3	2	2	3	3
12	1	1	1	1	1	1	1
13	1	2	2	2	1	2	2
14	2	2	2	1	2	2	2
15	1	2	2	1	1	2	2
16	2	3	3	2	2	2	3
17	2	1	3	2	2	1	1
18	1	1	1	1	1	1	1
19	1	2	2	1	1	2	2
20	2	3	3	2	2	3	1
21	2	1	1	2	2	1	1
22	1	2	1	1	1	2	2
23	1	1	1	1	1	2	1

Figure (2): Data

So what is the difference between this work and other related works? The most important and vital difference is about the kind of the data, first of all we collected data from various places and different situations so that the results can be more authenticated while in the related works most of the date had been collected from one place or one source . Secondly the period of the data it covers about fifteen years of events and experience so that the results could be more effective in the results of the prediction process, that wasn't the case in the related works where the data was about five years as most. In the similar related works the data was from war games and from training exercises but here we put real battles' data so the results could be more real. So we think that this work is unique because of its unique data so the result would be more authenticated.

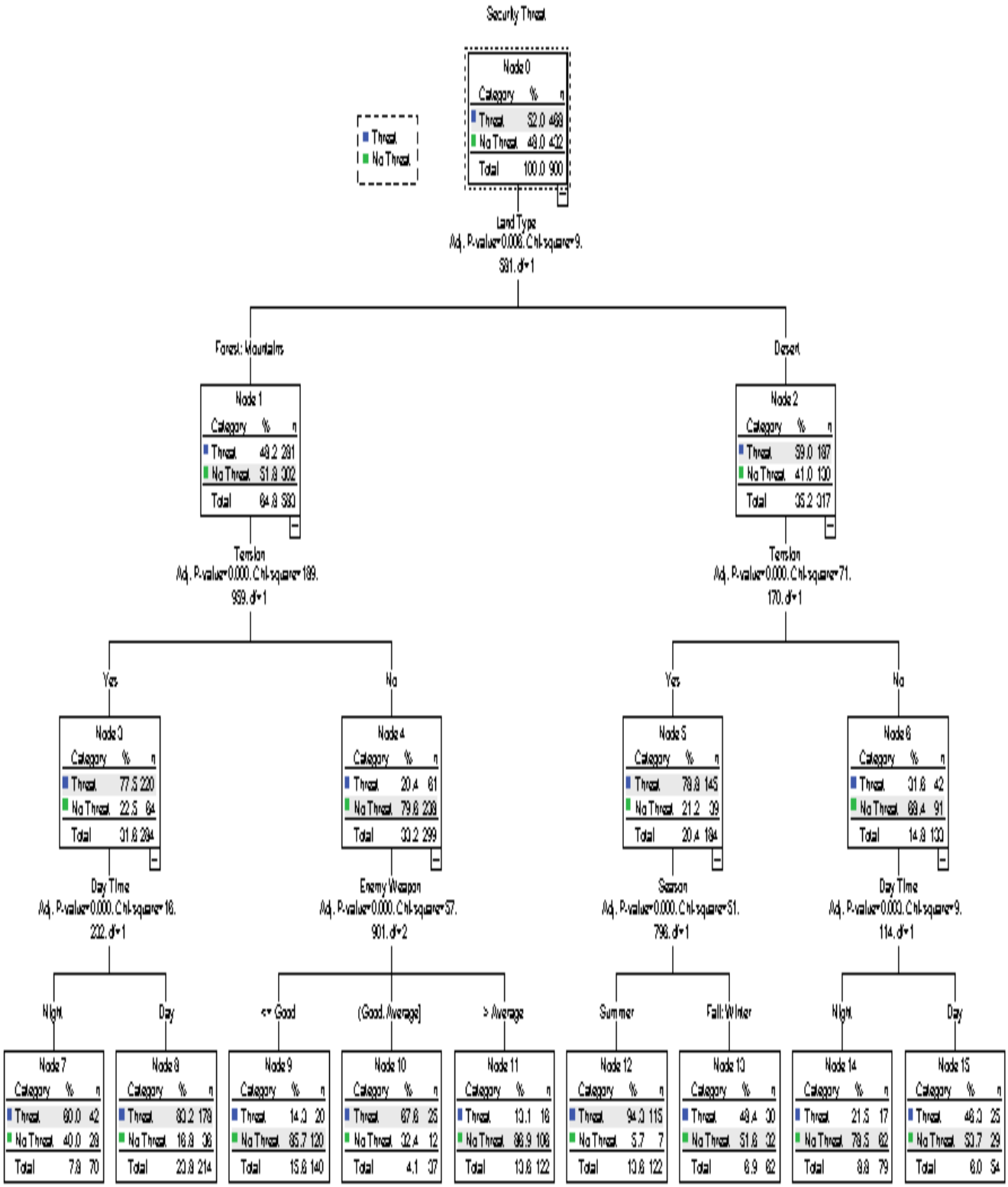


Figure (3): Model

Specifications	Dependent Variable	CHAID	
	Dependent Variable	Security Threat	
	Independent Variables	Day Time, Season, Tension, Enemy Weapon, Land Type	
	Validation	None	
	Maximum Tree Depth		3
	Minimum Cases in Parent Node		20
	Minimum Cases in Child Node		10
Results	Independent Variables Included	Land Type, Tension, Day Time, Enemy Weapon, Season	
	Number of Nodes		16
	Number of Terminal Nodes		9
	Depth		3

Figure (4): Model Summary

7. Conclusion and future work

This paper is expected to contribute in the process of decision making in military operations , so commanders can take their precautions according to the prediction results , whether there will be a threat or not . The research can be evaluated even more to enhance the accuracy of prediction using decision tree or any other suitable tool.

References

[1] M Millan, (2010) Entin and Serfaty. Defining and Assessing Expertise in a Complex Domain - Measures based on Theory. 37th Annual Meeting of the Human Factors and Ergonomie Society.

[2] J Gutierrez**,** G Leroy **)**2013) Claremont Graduate University Using Decision Trees to Predict Crime Reporting.

[3] J Farrar (2013). Decision Tree Algorithms Predict the Diagnosis and Outcome of Dengue Fever in the Early Phase of Illness.

[4] O Gynecol**) ,**2010) Predicting cesarean delivery with decision tree models, Pennsylvania

[5] E Philip. (2012). Mellers Intelligent Management of Intelligence Agencies Beyond Accountability Ping-Pong, University of Pennsylvania.

[6] A Falleson, J Jon. (2011). Overview of ArmyTactical Planning Performance Research. Technica Report 984. (Alexandria, VA: U.S. Army Research Institute for Behavioral and Social Science.

[7] A Michel, G Rex (2012). Historical Development of the Estimate of the Situation, U.S. Army Research Institute Research Report 1577. Fort Leavenworth KS: Fort Leavenworth Field Unit.

[8] G Padraic (2011). Neville SAS Institute Inc. Decision Trees for Predictive Modeling August.

[9] A Smith, B Kevin (2013). The Calculus of War: The Role and Use of Quantitative Decision Aids at the Tactical Level of War. MMAS thesis. Fort Leavenworth KS: U.S. Army Command and General Staff College.

[10] Y Fan , Y Qin,(2013) Application of ID3 Algorithm in the Analysis of Satellite Telemetry Data, School of Computer Science and Engineering, Xi'an Technological University, Xi'an 710021, China .

[11] W KOK, A MAJ. (2010). SINGAPORE ARMED FORCES B Sc (Hons), Imperial College, University of London, M Sc (Industrial Engr), National University of Singapore...

[12] B Long, L Clyde (2010). Synchronization of Combat Power at the Task Force Level: Defining a Planning Methodology. MMAS thesis. Fort Leavenworth KS.

[13] V Vijay (2011). Fulton Chair Professor Arizona State University PSERC Seminar.

[14] G Klein. (2010). "Strategies of Decision Making," Military Review.

[15] A John (2013). Tactical Decision Making: A Proposed Evaluation Criteria Model for the Infantry Battalion's Tactical Estimate during Offensive Operations. MMAS thesis. Fort Leavenworth KS: U.S. Army Command and General Staff College.

[16] J Petersen. (2011) "Forecasting: It's Not Possible," Defense Intelligence, (The Joint Military Intelligence College Foundation, VA), 37-45.

[17] IBM SPSS Decision Trees 21, User's Guide, 2012.

[18] A Yasir. (2013) Department of Computer and Information Sciences Karary University Biologically.

[19] National security policy, securing the Gains of Democracy, republic of Philipines, (2011).