EFFECT OF PESTICIDES ON HUMAN LIFE THROUGH VISUAL DATA MINING

[1]RABIA IMTIAZ, [2]Dr. MALIK SIKANDAR HAYAT KHIYAL, [3]DR. SHAHID KHALIL, [4]AIHAB KHAN, [5]Dr. AHSAN ABDULLAH

[1]Graduate Department of Software Engineering, Fatima Jinnah Women University, The Mall, Rawalpindi.
[2]Prof. & Chairperson Department of Computer Science and Software Engineering, Fatima Jinnah Women University, The Mall, Rawalpindi.
[3]Department of Mechanical Engineering, University of Engineering, Texila.
[4]Asstt. Prof. Department of Software Engineering, Fatima Jinnah Women University, The Mall, Rawalpindi.
[5]Prof. Department of Computer Science, Foundation University, Lalazar, Rawalpindi.

ABSTRACT

The wide spread of pro-pesticide in order to increase the yield of crops in Pakistan produces harmful effects on farmers as well as on economic, ecological and communal sector. This paper presents a new technique to analyze the agricultural data. This technique is based on visual data mining for gathering, organizing and examining the agricultural data. The main purpose of this technique is to put the cause of pesticide on human life. Chernoff faces are used to deduce the data in understandable form and the clustering of agricultural data is performed by Chernoff faces method. This paper explores the concept of multidimensional data. Experimental results have shown that by the use of multidimensional data display, integration and presentation have been enhanced.

Keywords*: Chernoff faces, Pesticide toxicity, Visual data mining,*

1. INTRODUCTION

Agriculture is the mainstay of Pakistan keeping in view its agrarian structure; it accounts for more than 24% of our GDP and 40% of our land mass used for cultivation is a means of income for the majority of our people. Thus, agricultural growth can help alleviate poverty and uplift our economy, thereby proving beneficial for everyone. As farming techniques are becoming more complex due to large-scale crop production, the need for timely information is also gaining more importance. This has led to the merging of two sciences viz. agriculture and Information Technology (IT), thus resulting in a unique field i.e. Agro-Informatics. Information Technology (IT) is a discipline that is employed for problem solving through those techniques and processes that have been tried and tested. It is a hybrid of different sciences comprising Agri-Sciences, Statistics, Computer Sciences, remote sensing, geographical information systems, etc. The importance of IT in the agricultural field can be determined by the farmers' reliance on modern IT for decision making processes and for gathering and recording data required for problem solving. This makes IT essential for our agricultural progress and consequently for improving the entire economic outlook. The adverse effects of pesticides on human life may leads to coma, convolution and even death. Toxicity is divided into four types, toxicity levels are defined on the basis of its LD 50 values. Furthermore, experience has shown that toxicity data obtained from a number of animal species can be useful in predicting human toxicity, while data obtained from single specie may be inaccurate. A list of classes of pesticides and their toxicity levels are given in Table 1.

Table 1: Toxicity Levels of Pesticide Classes

Pesticide Class	**Toxicity To Human**
Nicotinoid	Low
Pyrethroid	Low
Organophosphate	Medium
Pyrethroid + Organophosphate	Very High
Carbamate	Very Low

Excessive usage of pesticides is harmful for the environment and hazardous to human health. Pesticides effects can be reduced by looking for the conditions in which the usage is optimum and trying to dig out for the circumstances. In this study we have shown how visual data mining can be successfully applied for this purpose**.**

Rest of the paper is organized as follow; Section 2 gives background of our work, Section 3 presents a brief review of related work, Section 4 and 5 describe structures and working of Chernoff faces Algorithm, Chernoff faces application and discussion on results are presented in Section 6 and Section 7, while conclusions are summed up in Section 8.

2. WORKING SCENARIO

To learn from the past one needs a detailed record of the past. This data is obtained from the Center for Agro informatics Research (C@IR). As a pilot project they implemented a warehouse, the warehouse was implemented after digitization, cleansing and integration of data generated by multiple resources. The data used for scenario is only for District Multan, which are the thirty-four districts of Province of Punjab.

3. RELATED WORK

A study conducted by Ahsan Abdullah, et al., 2003 on Learning Dynamics of Pesticides Abuse through Data Mining reported that excessive use of pesticides is harming the farmers with adverse financial, environmental and social impacts. Unsupervised clustering of agricultural data through Recursive Noise Removal (RNR) heuristic has been performed to identify the pesticide dynamics and reasons for pesticides usage abuse [1].Another study by Volker Markl, et. al, 1999 introduces the idea of using Multidimensional hierarchical clustering (MHC) of OLAP to speed up the aggregation queries and improving the performance and cost measures in data warehousing. An encoding scheme has been defined for hierarchical dimensions that enable clustering of data with respect to multiple hierarchical dimensions [2].Alan Dix and Geoffrey Ellis emphasis the established techniques to allow visual effects such as shapes, colour, and Chernoff faces [Chernoff 1973] [3]. A research paper titled ‘ Am Emperical Study into the use of Cherno_Information for Robust, Distributed Fusion of Gaussian Mixture Models’ by Simon J. Julier, uses Chern off information for the distributed fusion of Gaussian Mixture Models[4]. Michael D. Lee et.al described that Chernoff faces and star glyphs, represent objects using simple icon-like displays. All features were present on the Chernoff faces, this means, the presence or absence of, say, a mouth, but by extremes in its length or curvature corresponding the software’s default values. The visualization represents a group of twenty animals which possesses one or more of a selection of physical features [5]. Robert P. Bosch, Jr. discussed the Chernoff faces in the visualization of the recovery algorithm, employing the physical arrangement of nodes in the machine to create an array of Chernoff faces [6].Ethan Perry and Judith Donath used the human form to visualize historical information about participants in online spaces, They created an approach that is evocative and raises some intriguing questions [7]. The PMRA uses a well defined decision framework. An assessment of risks to human health is to define the nature of the risk and to provide a measure and magnitude of the risk associated with a defined exposure. The PMRA decision making is designed to protect human health and the environment and only allow pesticides that provide value to users and the Canadian society to be registered [8]. Risk of these diseases in developing countries is high where pesticide is commonly use and also basic safety precautions are not adopted. The Food and Agriculture Organization of the United Nations (FAO recommends that World Health Organization Class Ia, Ib and, preferably II, pesticides should not be used in developing countries. However, such chemicals remain widely available in these countries. The Code of Conduct needs to be strengthened with regard to measures aimed at protecting human health and more effective monitoring of its implementation is required [9].

4. FRAMEWORK OVERVIEW

The model in fig 1 shows that the input is provided in the form of data tables. First of all it checks the availability of database. If there is no database exist it print the error. Otherwise select the table from the list. The selected columns display in the grid and the second grid display the data conversion of these selected columns. After reordering the dataset it convert the selected data into faces. The desired results of the system are shown in the form of human faces.

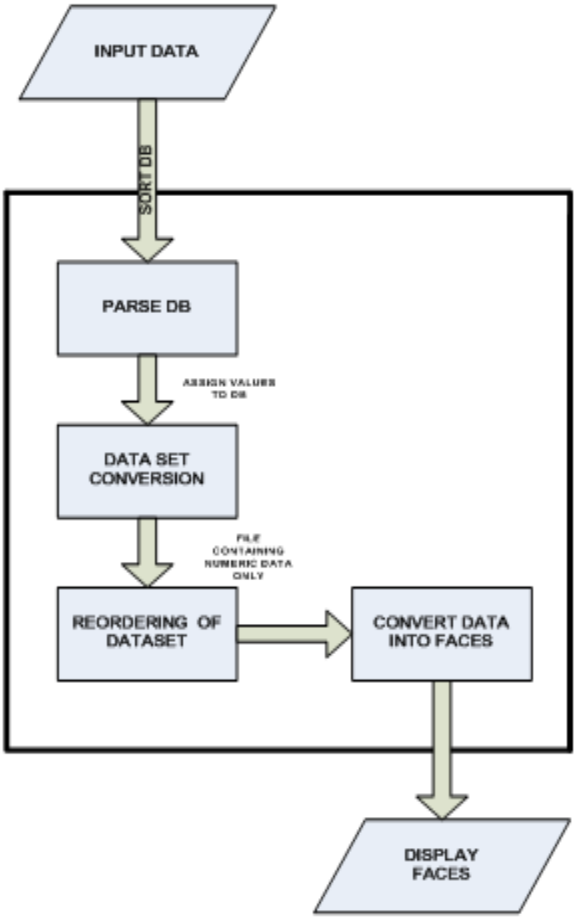


Fig 1: System Model

5. PRELIMINARIES

Humans are very good at perceiving facial expressions and at grouping facts by certain characteristics. Thus, if the data is interpreted in the form of facial expressions, understanding of the data would become very easy. 'Chernoff faces' is a useful icon based technique which uses features in cartoon-like human faces, each representing variables in order to depict multivariate data. Chernoff faces shows clustering on the basis of similar features.. Chernoff developed a unique graphical method to encode and display multivariate data by using iconic faces, each representing individual records. This method is different from other methods of multivariate data visualization, where work is done by reducing the information in the data set to a relatively few variables which results in the loss of original information, hiding of results and hiding of findings in the data. Figures 2 and 3 shows the geographical location of district Multan and also elaborate the geographical area included in this research.

Fig 2: Map of Pakistan [10]

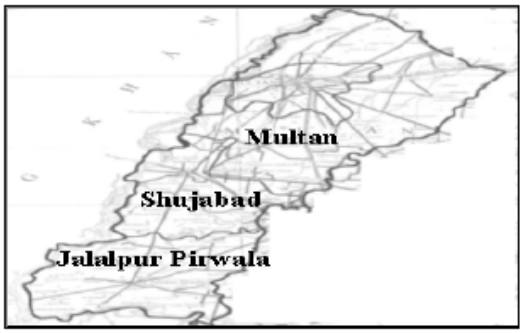
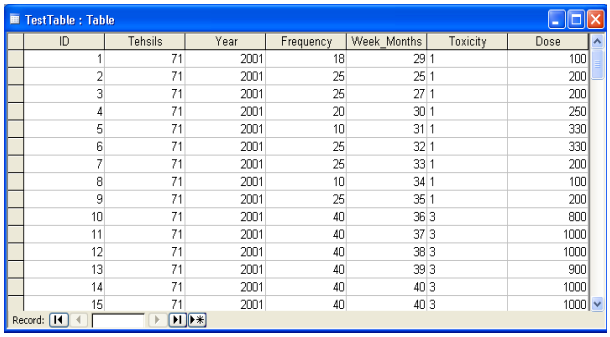


Fig 3: Area under Study [10]

6. TECHNIQUE

The algorithm proposed for visual data mining technique has the following steps.

- *Load the database on which the data mining algorithm is to be carried out*


TestTable : Table

ID	Tehsils	Year	Frequency	Week_Months	Toxicity	Dose
1	71	2001	18	29	1	100
2	71	2001	25	25	1	200
3	71	2001	25	27	1	200
4	71	2001	20	30	1	250
5	71	2001	10	31	1	330
6	71	2001	25	32	1	330
7	71	2001	25	33	1	200
8	71	2001	10	34	1	100
9	71	2001	25	35	1	200
10	71	2001	40	36	3	800
11	71	2001	40	37	3	1000
12	71	2001	40	38	3	1000
13	71	2001	40	39	3	900
14	71	2001	40	40	3	1000
15	71	2001	40	40	3	1000

Record:

Fig 4: Input Data

- *Select the number of columns from the data tables.*
- *The application will take data for each column sorted out and categorized for generating visual input values.*
- *Assign the shape values to the individual cell and display them in grid.*

- *Sort the column, and find the numbers of possible shapes assign to each column*

  Max value-Min value/3= Normalized value
  Value< Normalized value=1
  Value=Normalized value=2
  Value>Normalized value=3

Converted Dataset

	Tehsils	Year	Frequency
▶	2	1	2
	2	1	1
	2	1	1
	2	1	1
	2	1	1
	2	1	1
	2	1	1
	2	1	2
	2	1	1

Fig 5: Normalized Dataset

- *Choose the column/field in which you want to visualize the data. It will calculate position for each record depending upon its dataset value.*
- *Forward the dataset for display output.*
- *The output will be in the form of human faces which represent three types of effects: Happy, Normal and Sad.*
- *Get each element from dataset and after assigning relevant color to it, draw faces on the screen randomly.*
- *The shape will have similar effects or similar values will be drawn in the form of clusters.*
- *This would be followed by the transformation of data and scaling of the parameter values, which would result in complete algorithmic faces presentation.*
- *The results will be found on the basis of happy face, sad face and normal face.*

7. EXPERIMENTAL RESULTS

The parameters used to find the affects of pesticides on human life are Tehsils, Year, Toxicity, Pesticides Frequency, Months and dose used in District Multan. Six cases have been considered and used for the analysis based on different selection criteria, like geography, time and vice versa. The output shows happy, normal and sad faces that represent different effects of pesticide toxicity levels on human life in different tehsils of District Multan. Although we can perform analysis using conventional charts and graphs, but they show a more abstract level of details of the data, using sums and aggregates, in which we can not view all records individually. However, by using Chernoff faces technique we can do the in-depth analysis by viewing each record in the form of face and outlier can be detected. Furthermore, by using this technique similar records forms a cluster and their distance shows the level of similarity between them

There are six different types of test cases that show different results

1) The first case represent the tehsils of district multan by selecting the parameter face. Similar group of faces represent the tehsil.
2) The second test case represents the number of years. In this case the selected criteria are eyes.
3) The third case represent the frequency of pesticides and the selected criteria in this case is nose
4) The fourth case represents the number of months by using the selection criteria eye brows.
5) The fifth case represents the toxicity levels by selecting the criteria mouth. Similar mouths represents different toxicity levels
6) The sixth case represents the pesticides dosage used by selecting the criteria of pupil position.

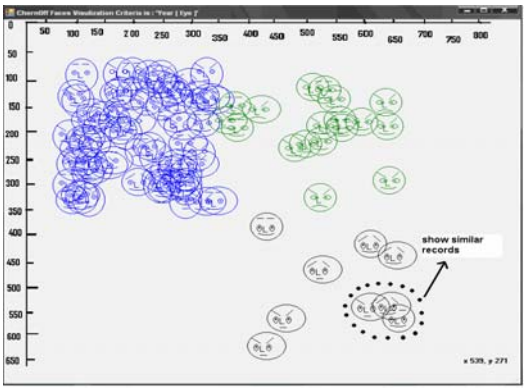


Fig 6: Results shows similar records

Analyses of the first and third clusters show that in the years 2001 and 2003, the maximum numbers of happy and normal faces represent the low and medium toxic effects of pesticides on human life. In comparison to this, in Sujahabad we see happy and sad faces showing the low and

high toxic affects of pesticides on human life in the year 2002. Similar records form a cluster and their distance shows the level of similarity between them, as shown in fig 6.

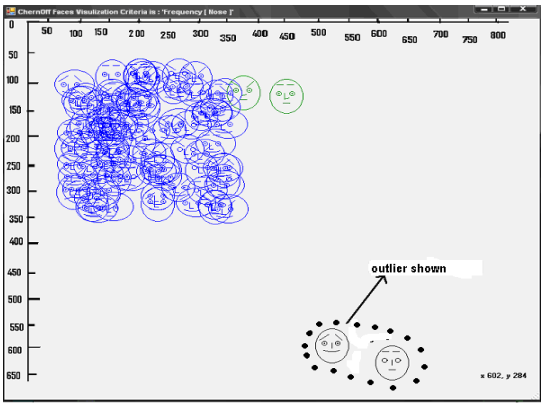


Fig 7: Results shows outlier

Analysis of first cluster shows that there has been maximum frequency of pesticides in the year 2001 in comparison to the frequency observed in the years 2002 and 2003, respectively. We can do the in-depth analysis by viewing each record in the form of face and outlier can be detected, as shown in fig 7.

6. SUMMARIZED RESULTS

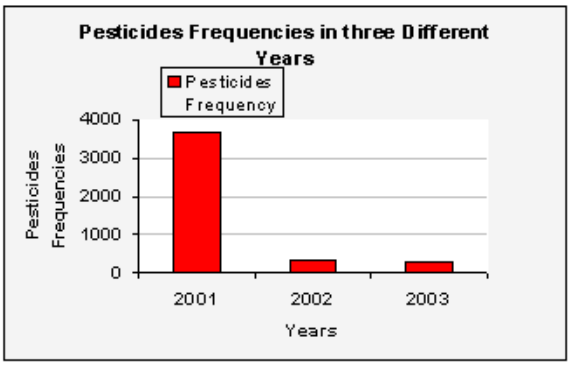


Fig 8: Pesticides frequency in three years

The fig 8 shows the maximum pesticide frequency in 2001 in comparison to the frequency observed in the years 2002 and 2003, respectively.

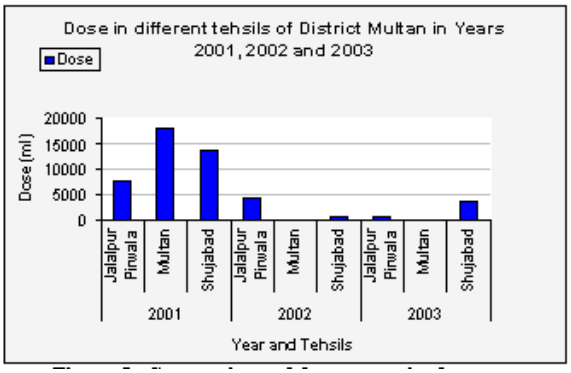


Fig 9: Comparison of dose usage in three years

The fig 9 shows the maximum pesticide dose usage in 2001 in comparison to observed in the years 2002 and 2003, respectively.

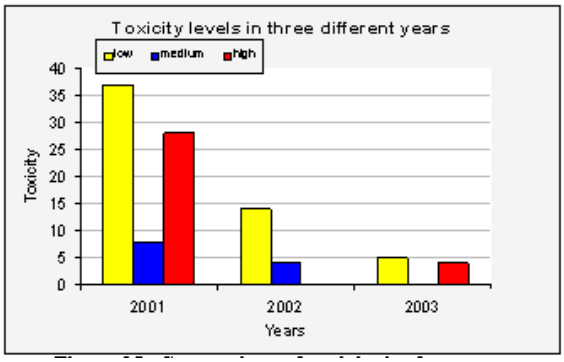


Fig 10: Comparison of toxicity in three years

The fig 10 shows the low, medium and high toxic affects on human life in the year 2001,2002 and 2003.

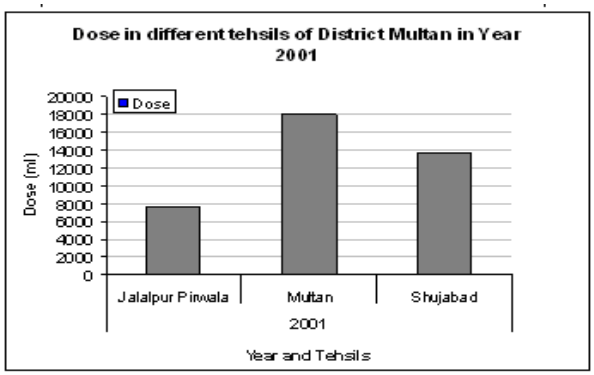


Fig 11: Dose in different tehsils of Multan

The fig 11 shows the maximum amount of pesticide dose used in Multan in 2001 in comparison to the dose used in other two tehsils.

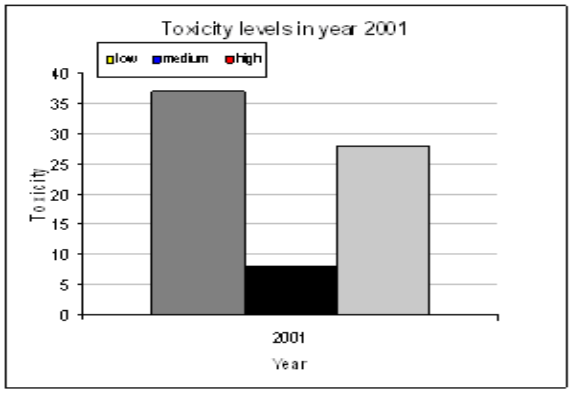


Fig 12: Comparison of Different Toxicity levels in Year 2001

The fig 12 shows the low, medium and high toxic affects on human life in the year 2001.

8. CONCLUSION AND FUTURE WORK

The results concluded from the above study are shown in the form of human faces by using pesticide data. Pesticide data exists in a complex form which is difficult to be understood by simple users. To facilitate the decision making process of farmers, the data is normalized. For the normalization, the complex data has been preprocessed and converted into simple form. Clustering is performed on the basis of selection criteria of parameters. Effect of pesticides on humans can't be directly checked because of the poisonous nature of pesticides, therefore the usage of pesticides on cotton crop has been taken into consideration for this purpose. The COF Clustering Tool can not only be used for pesticide data, but also possesses the flexibility to deal with any numeric data.

This tool can use to analyze only the numeric values and takes only six parameters. So far the future work we could solve this problem by using maximum 18 parameters and enhance the functionality so as to handle textual data.

REFERENCES:

[1] Abdullah, A. and Brobst, S. (2003): Clustering by recursive noise removal. Proc. Atlantic Symposium on Computational Biology and Genome Informatics, USA, pp. 973-977.

[2] Markl et al (1999): Improving OLAP Performance by Multidimensional Hierarchical Clustering. Bayerisches Forschungszentrum für Wissensbasierte Systeme Orleansstraße 34, 81667 München, Germany

[3] Alan Dix and Geoffrey Ellis (0000): starting simple adding value to static visualization through simple interaction AVI98 paper .

[4] Simon J. Julier (0000): An Empirical Study into the Use of Cherno Information for Robust, Distributed Fusion of Gaussian Mixture Models, ITT AES / Naval Research Laboratory 4555 Overlook Avenue SW Washington, DC 20375, USA

[5] Michael D. Lee et.al (0000): An Empirical Evaluation of Chernoff Faces, Star Glyphs, Spatial Visualizations for Binary Data .

[6] Robert P. Bosch, Jr.(0000): Distributed Algorithm Visualization.

[7] Ethan Perry and Judith Donath (0000): Anthropomorphic Visualization: A New Approach For Depicting Participants in Online Spaces, MIT Media Laboratory Cambridge, MA 02139 USA

[8] A Decision Framework for Risk Assessment and Risk Management in the Pest Management Regulatory Agency, Pest Management and Information Division,Pest Management Regulatory Agency. Health Canada A.L. 6602A Ottawa, Ontario K1A 0K9

[9] A report on Health threat posed by pesticides in developed countries by the Environmental Justice Foundation .

[10] A. Abdullah et.al (2004) "Learning Dynamics of Pesticide Abuse through Data Mining" in proceedings of Australasian Workshop on Data Mining and Web Intelligence, Dunedin, New Zealand