

EXPLORATORY DATA MINING OF SOME PLANT SPECIES AGAINST *Aedes aegypti*

Rajmohan Devadass¹, Haldurai Lingaraj² and Logankumar Kandhasamy³

^{1,3}PG and Research Department of Zoology

²Department of Computer Science (PG)

^{1,2,3} Kongunadu Arts and Science College, Coimbatore, Tamilnadu, India



Date Received:

05-Aug-2014

Date of Accepted:

16-Aug-2014

Date Published:

29-Aug-2014

Abstract:

Biological databases contain a wide variety of data types, often with rich relational structure. On sequence multi-relational data mining techniques frequently are applied to biological data. Extracts of the different plant species were evaluated for the egg hatchability, larvicidal and pupicidal activity of mosquito, *Aedes aegypti* under the room temperature in the laboratory. Dosage value as expressed in ppm was 10 to 400 for *Aedes aegypti*. A relationship was observed between the plant extract doses and percentage mortality. The percentage of egg hatchability, larval and pupal mortality was found to be increased with increase in the dosage. Based on the probit analysis, the L_{C50} value of egg, instars and pupae hence assumed.

Keywords: Data mining, *Aedes aegypti*, ppm, L_{C50} , Prediction.

Introduction

Data mining is not a solution in itself, but a tool that supports the decision making process in finding a solution and that provides the information required to solve a problem. Data mining helps the analyst to discover patterns and relationships contained within the data.

Knowledge discovery in databases is well-defined process consisting of several distinct steps. Data mining is the core step, which results in the discovery of hidden but useful knowledge from massive databases. A formal definition of Knowledge discovery in databases quotes "Data mining is the non trivial extraction of implicit previously unknown and potentially useful information about data" (Frawley and Piatetsky-Shapiro, 1996).

Vector control is a serious concern in developing countries like India due to lack of general awareness, development of resistance and socio-economic reason. The role of mosquitoes is becoming increasingly important in the recent years because of change in ecology caused by human intervention.

Mosquitoes constitute a major public health menace as vectors of serious human diseases (El Hag *et al.*, 1999). Of the various mosquito spread diseases, dengue fever transmitted by *Aedes aegypti* is dangerous and is spreading into dengue shock syndrome have taken epidemic form and have been reported from Tamil Nadu, West Bengal, Uttar Pradesh, Gujarat and Delhi (Kebra *et al.*, 1992). In recent years, scientists try a variety of botanical derivatives to eradicate many harmful insect pests including mosquitoes.

For the present study, the data mining of plant species were screened against egg hatchability, larval and pupal mortality of the mosquito *Aedes aegypti*.

METHODOLOGY

In this study ten plant species were taken such as *Tridax procumbens*, *Annona squamosa*, *Tagetes erecta*, *Acorus calamus*, *Casearia esculenta*, *Ricinus communis*, *Pongamia pinnata*, *Vitex negundo*, *Lantana camara* and *Ageratum conyzoides* and their extracts have been tested against *Aedes aegypti* with various dosage (ppm). Reading were taken at regular intervals and are stored in the database. To retrieve a particular data based on some criteria it will be a difficult task. Hence Data Mining as an analytic process designed to explore large amounts of data in search for consistent patterns and systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data.

STATISTICS FOR PREDICTION

The term "prediction" is used for a variety of types of analysis that may elsewhere be more precisely called regression. We have done so in order to simplify some of the concepts and to emphasize the common and most important aspects of predictive modeling.

COMPUTATIONAL EDA TECHNIQUES

Computational exploratory data analysis methods include both simple basic statistics and more advanced, designated multivariate exploratory techniques designed to identify patterns in multivariate data sets.

The basic statistical exploratory methods include such techniques as examining distributions of variables (e.g., to identify highly skewed or non-normal, such as bi-modal patterns), reviewing large correlation matrices for coefficients that meet certain thresholds or examining multi-way frequency tables

Multivariate exploratory techniques designed specifically to identify patterns in multivariate (or univariate, such as sequences of measurements) data sets. The exploration of data can only serve as the first stage of data analysis and its results can be treated as tentative at best as long as they are not confirmed, e.g., cross validated, using a different data set (or an independent subset). If the result of the exploratory stage suggests a particular model, then its validity can be verified by applying it to a new data set and testing its fit (e.g., testing its *predictive validity*). Case selection conditions can be used to quickly define subsets of data (e.g., for estimation and verification), and for testing the robustness of results.

RESULT

The above graphs convey lot of information based on database for a particular predictor or data column. By looking at the graph it is also possible to build an intuition about other important factors. Such as the mortality rate with maximum and minimum ppm. Table.1 shows the effect of extracts of ten plants species against the developmental stages of the mosquito, *Aedes aegypti*. Fig.1 reveals plant wise records with lowest ppm which shows highest mortality rate among ten plants. Fig.2 shows plant wise records with highest ppm which shows highest mortality rate among ten plants. Fig.3 lists stage wise records with lowest ppm which shows highest mortality rate among ten plants. Fig.4 shows stage wise records with highest ppm which shows highest mortality rate among ten plants. From the results we can conclude *Tridax procumbens* shows high efficacy on *Aedes aegypti* (fig.3) and *Ageratum conyzoides* shows minimum efficacy on *Aedes aegypti* (fig.4).

CONCLUSION AND CHALLENGES

Biology and data mining are developing as interdisciplinary science. Data mining approaches seem ideally suited for Life sciences, since biology is data-rich but lacks a comprehensive theory of life's organization at the molecular level. However, data mining in biology is hampered by many facets of biological databases, including their size, number, diversity and the lack of a standard ontology to aid the querying of them as well as the heterogeneous data of the quality and provenance information they contain. Another problem is the range of levels the domains of expertise present amongst potential users, so it can be difficult for the database curators to provide access mechanism appropriate to all. The integration of biological databases is also a problem. Data mining and bioinformatics are fast growing research area today. It is important to examine what are the important research issues in life science and develop new data mining methods for scalable and effective analysis.

REFERENCES

- Frawley and Piatetsky-Shapiro, 1996.** Knowledge Discovery in Databases:An Overview. The AAAI/MIT Press, Menlo Park, C.A.
- El Hag EA, Nadi AH and Zaitoon AA. 1999.** Toxic and growth retarding effects of three plant extracts on *Culex pipiens* larvae (Diptera; Culicidae) Phytother. Res. 13: 388-392.
- Kebra SK, Verma IC, Arora NK, Jain Y and Kalra V. 1992.** Dengue Haemorrhagic fever in children in Delhi. Bull. WHO, 70(1): 105-108.

Table : 1 Effect of extracts of ten plants species against the developmental stages of the mosquito, *Aedes aegypti*

Stages	LC50 (ppm)									
	Tridax procumbens	Annona squamosa	Tagetes erecta	Acorus calamus	Casearia esculenta	Ricinus communis	Pongamia pinnata	Vitex negundo	Lantana camara	Ageratum conyzoides
Egg	39.06	58.93	74.49	88.43	141.78	202.54	232.46	258.27	265.61	362.04
I Instar	16.52	18.73	18.41	31.73	68.25	137.95	123.69	127.54	177.87	221.19
II Instar	19.06	24.02	31.35	37.15	78.74	171.28	163.19	149.68	198.55	252.37
III Instar	33.91	32.54	35.06	68.84	99.58	173.07	209.38	182.86	216.17	307.82
IV Instar	39.45	44.16	48.62	72.55	122.80	198.84	217.64	198.40	243.62	324.35
Pupa	51.08	54.49	81.94	84.13	125.52	249.60	275.47	226.97	270.48	372.94

Fig 1. Records with lowest ppm which shows highest mortality rate among 10 plants. (plant wise)

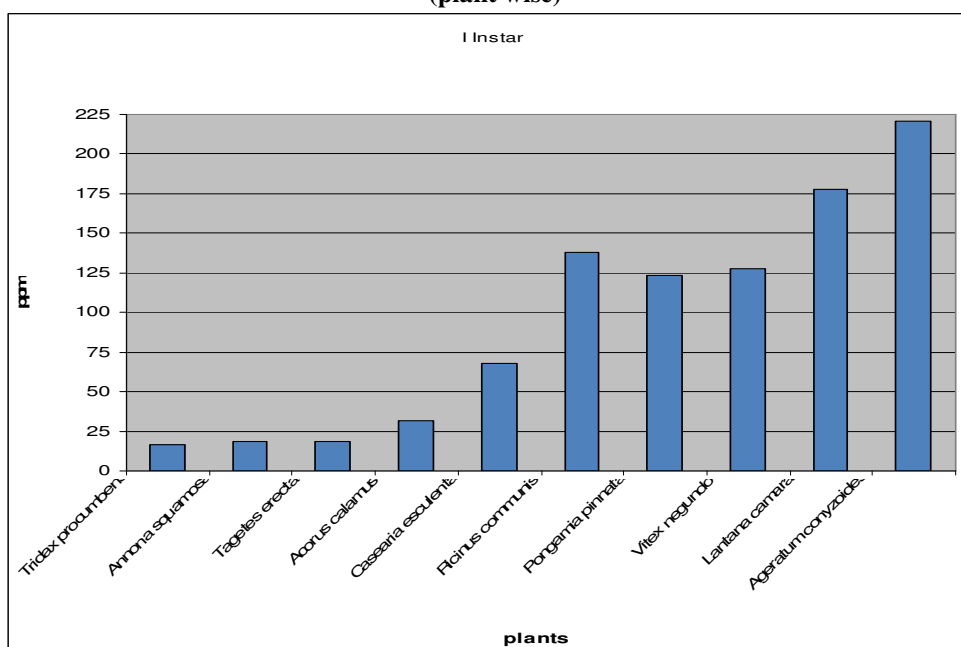


Fig 2. Records with highest ppm which shows highest mortality rate among 10 plants.

(plant wise)

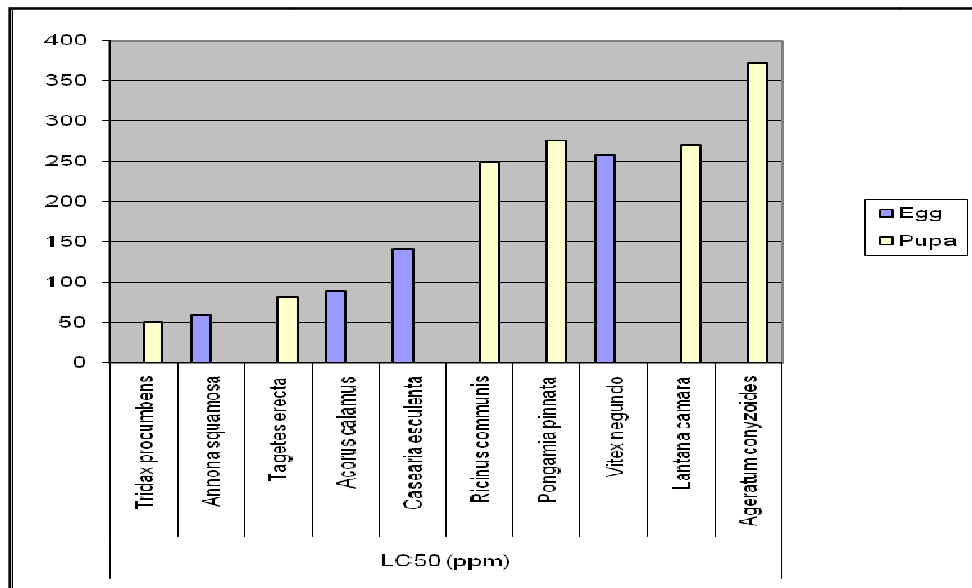


Fig 3. Records with lowest ppm which shows highest mortality rate among 10 plants.
(stage wise)

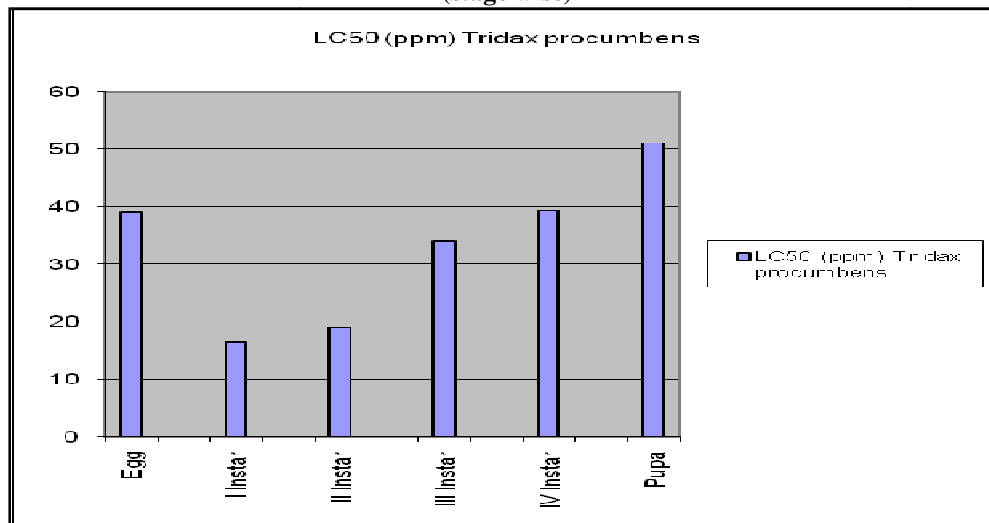


Fig 4. Records with highest ppm which shows highest mortality rate among 10 plants.

(stage wise)

