ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

An Implementation of Dengue Fever Disease Spread Using Informatica Tool with Special Reference to Dharmapuri District

S. Stany Leena Princy[1], A. Muruganandam[2]

Research Scholar, Department of Computer Science, Don Bosco College, Periyar University, Salem, Tamil Nadu, India[1]

Research Supervisor, Department of Computer Science, Don Bosco College, Periyar University, Salem, Tamil Nadu, India[2]

ABSTRACT: Dengue virus belongs to the family of Flaviviridae, having four serotypes (DENV 1–4). It spread by the bite of infected Aedes mosquitoes. Dengue fever is caused by any one of four dengue viruses spread by mosquitoes that succeed in and near human lodgings. When a mosquito bites a person impure with a dengue virus, the virus enters the mosquito. When the infected mosquito then bites another person, the virus enters that person's bloodstream. Dengue can vary from a simple fever to severe fatal Dengue Haemorrhagic Fever / Dengue Shock Syndrome (DHF/DSS). Data mining is a prominent technology which is used by health sectors for identifying diseases like dengue, diabetes and cancer**.** Data sets were collected from Dharmapuri health center. Our analysis shows, out of the 100 serum samples, which include 43 female and 47 male, 30 percent a positive result for dengue-virus and 70 percent was negative. Informatica tool used in this paper is being for purpose of analysis of the provided data set and to detect the dengue affected patients for early and good recovery.

KEYWORDS: Dengue, Flaviviridae, DHF, Dengue Shock Syndrome and Informatica**.**

I. INTRODUCTION

Dengue Fever (DF) is a highly Infectious Disease which spreads very easily. It is very essential to know about the disease before it spread from one State to another. It is an unpleasant sudden illness of a week duration (sometimes with two weeks) with two or more of the following symptoms: Headache, Body pain, abdominal pain, continuous vomiting, fever, reduction in WBC platelets and nose bleeding and severe bone pain. DF is usually mild in children whereas it is severe in adults. Dengue Haemorrhagic Fever (DHF) is known for the following symptoms, Positive tourniquet test, Small red or purple colored spots, purpura that means spontaneous bleeding under the skin surface, bleeding from gums. Reduction in platelets one lakh/cu.mm or less and Evidence of plasma leakage due to increased blood vessel wall for the flow of small molecules. Dengue Shock Syndrome (DSS) all the above specified symptoms plus the signs of weak pulse, low pressure, hypotension for age, fever, cold, restlessness. The above descriptions are enough for Doctors to determine signs of Dengue and to treat the same. However further tests are to be conducted to classify whether it is DF/DHF/DSS using the above criteria. An early diagnosis can help in quick recovery. For better and accurate result Informatica techniques are used for separating a dengue affected patient from a normal patient [1].

II. RELATED WORK

The incidence of Dengue fever in India is increasing day by day and so it is very important to know more about the dangerous disease. Early detection and prevention helps quickly recovery. Global warming, population growth and rapid urbanization play an important role in spreading dengue. The author says that there are two ways to control dengue: control of Aedes mosquito which spreads dengue virus, and the vaccines to control the dengue virus infection [2].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

A Study of Clinical Profile of Dengue Fever in a Tertiary Care Teaching Hospital" refer to dengue as an international health concern very powerful and spreading widely in tropical and sub-tropical countries. The authors have presented a study on north Karnataka as it is largely affected by dengue virus. They conclude that dengue in the major cause of undifferentiated fever and is hardly recognized as a clinical entity by physicians [3].

A retrospective analysis "Dengue is one of the most important Arbovirus diseases seen in Southeast Asia and India. The authors have made an analysis of dengue with the referred samples of a three year period (2006-2008) at the department of virology, Chennai. They come to the conclusion that creating a general awareness among the public and through continuous vigilance from the health care physicians could go a long way in combating dengue [4].

Chikungunya virus is very similar to dengue virus it spreads very fast in tropical regions due to the presence of Aedes mosquitoes. The techniques of machine learning and algorithms of data mining play an important role in diagnosis and prognosis of several health diseases. The authors focus on comparing the chikungunya viral affected patients with dengue viral affected patients [5]. Dengue is a life threatening disease caused by female mosquito's called "Aedes", usually found in tropical regions. Officials and experts have, for a long time, been in search of details on dengue, so that they can accurately classify the dengue affected patients as they require on exclusive treatment. For the last few years Pakistan has been the target of dengue disease. The authors, in their search of dependable results have used classification techniques for evaluation of their work. [6]. "Analysis of significant factors for dengue infection prognosis using the self-organizing map" presents a new approach of using SOM map. This map has been used for visualizing and categorizing the dengue affected person from normal healthy subjects [7].

Dengue virus belongs to family Flaviviridae that has four serotypes spread by Aedes mosquitoes. The above 2.5 million people living in dengue threat regions for approximately. Dengue disease creates highly complex economic and ecologic problems in the society. The author explained the Extent of work done by various groups of scientist in this country [8].

Data mining is used widely by many organizations. Especially in health sector data mining is getting popular in the today world. This organization tool results in good decision making leads to effective treatment for patients so that they can receive a better and affordable service by health sector. Author used clustering and classification techniques of data mining to find the hidden patterns which will result in meaningful decision making [9].

III. METHODOLOGY

Informatica is a software development company found in 1993. The company headquarter is located in Redwood City, California. Gaurav Dhillon and Diaz Nesamoney is the co-founder of this company. The CEO of the company is Anil Chakravarthy. Informatica Power Center is widely used Extraction Transformation Loading ETL tools used in building enterprise data warehouses to store and prevent the data in data base. The Informatica Power Center aid in extracting data from its source, transforming it as per business requirements and loading into a target data warehouse. Extracting Transformation Loading can take data that is heterogeneous and make it homogeneous form. Without ETL it would be impossible to programmatically analyze heterogeneous data and derive business intelligence from it [10].

Extract the first phase of ETL, in which the data is collected from different sources and stored in a temporary storage where the removable of duplication is executed. During this process data are tested whether it has the expected values to store in data warehouse which is very essential if the data fails to satisfy then it may be removed or processed further to discover and remediate if possible. This process is done by using the validation rules. Transform- the second phase of ETL, in which the data processed to be valued and structured across all the data. The transformation phase include things like date formatting, resorting rows or columns of data, joining data from two values into one, or, conversely, splitting data from one value into two. The transformation goal is to make all the data are in uniform schema. Load- the third phase of ETL, in which the transformed data are moved permanently into the target database [11].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Once the data are loaded to the target the ETL process is complete, in order to make the data warehouse updated with the latest data ETL is processed regularly in many organization. Informatica is an enterprise tool which is mainly used for big data analyses. Many other data mining tools fail to work with big data, but Informatica is more accurate when processing more number of data. It can deal with lakhs and lakhs of data. The assets of Informatica are: Informatica jobs (sessions) can be arranged logically into worklets and workflows in folders. Job monitoring and recovery (Easy to monitor jobs using Informatica Workflow Monitor). Easier to identify and recover in case of failed jobs or slow running jobs. Ability to restart from failure row step. Informatica Market Place one stop shop for lots of tools and accelerators to make the SDLC faster, and improve application support [12].

Fig.1: has shown the process involved in the analysis of the data in Informatica tool. The source file used by us for importing data is in CSV format, the raw data is now trimmed, formatted, de duplicated and transformed to meaningful data with the help of expression transformation. The data is then processed in the router transformation rather than using a filter in which we can provide only one condition which drops the un-satisfied records. We have used router transformation in which we can give one or more condition which provides the ability to save unprocessed data in a different target for future use. This is done for avoiding the bottleneck. Mapping designer in Informatica tool is used to engender the map based on the relation. Transformations are used to differentiate dengue patients from healthy objects by using Mapping designer.

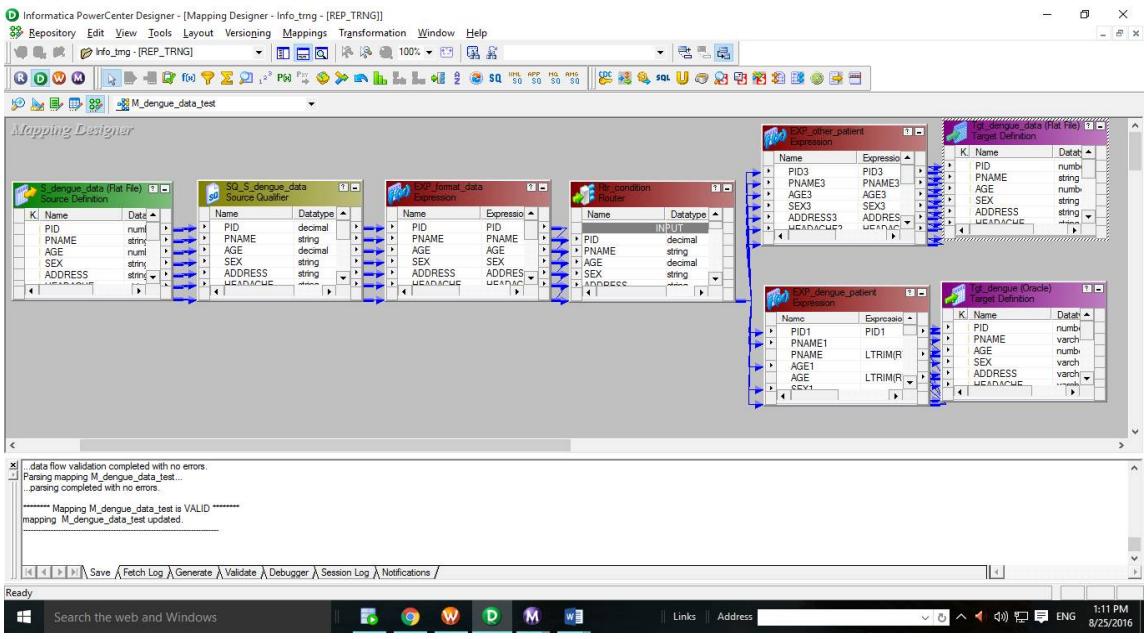


Fig.1: Mapping Process

A. 3.1 Data set

Database of 100 datasets (30 dengue patients and 70 healthy subjects) was obtained from Dharmapuri Health Center. 7 symptoms and signs were investigated on the day of elevated temperature of fever. The data comprised headache, body pain, abdominal pain, continuous vomiting, fever, bleeding tendency and reduction in WBC platelets. We have differentiated the dengue patients from the healthy subjects through use of the transformation of Informatica. This study was conducted using Informatica Techniques (number of transformations). Table 1 below shows, the different data is used for the diagnose to find the presence of dengue.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Table 1. The chunk of Data set

PID	AGE	SEX	HEADACHE	BODY PAIN	ABDOMINAL PAIN	VOMITING	FEVER	WBC COUNT	BLEEDING	RESULT
1	45	M	yes	yes	yes	yes	yes	200000	yes	+ve
2	50	M	yes	no	yes	no	yes	500000	no	-ve
3	35	M	no	yes	yes	no	yes	600000	no	-ve
4	55	M	yes	yes	yes	yes	yes	100000	yes	+ve
5	60	M	yes	yes	no	no	yes	500000	no	-ve
6	45	M	yes	yes	no	no	yes	700000	no	-ve
7	20	F	yes	yes	yes	yes	yes	300000	yes	+ve
8	52	M	no	yes	no	yes	yes	550000	no	-ve
9	30	M	yes	yes	yes	yes	yes	320000	yes	+ve
10	48	M	no	no	no	yes	yes	800000	no	-ve
11	39	M	yes	no	no	yes	yes	850000	no	-ve
12	28	F	no	no	no	yes	yes	500000	no	-ve
13	65	F	no	yes	no	yes	yes	650000	no	-ve
14	58	F	yes	no	no	no	no	800000	no	-ve
15	63	F	yes	yes	no	no	yes	750000	no	-ve
16	3	M	yes	yes	yes	yes	yes	250000	yes	+ve
17	8	F	no	no	yes	yes	yes	900000	no	-ve
18	10	F	yes	no	yes	yes	yes	660000	no	-ve
19	5	F	no	no	yes	yes	yes	780000	no	-ve
20	4	M	yes	yes	yes	yes	yes	390000	yes	+ve
21	8	F	yes	no	no	no	no	980000	no	-ve
22	9	M	yes	yes	yes	yes	yes	156000	yes	+ve

The sequence of diagnosis performed is based on the actual symptoms, which in turn, is based on the results identified for the provided sample as affected by dengue virus. We have taken one serum for our test to determine if it is affected by dengue. The serum goes through a variety of tests satisfying the below condition and will be termed as dengue affected serum. When anyone of the conditions fails, it can be termed as non – denguesample. Fig. 2 shows the flow of different tests taken and the actual results.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

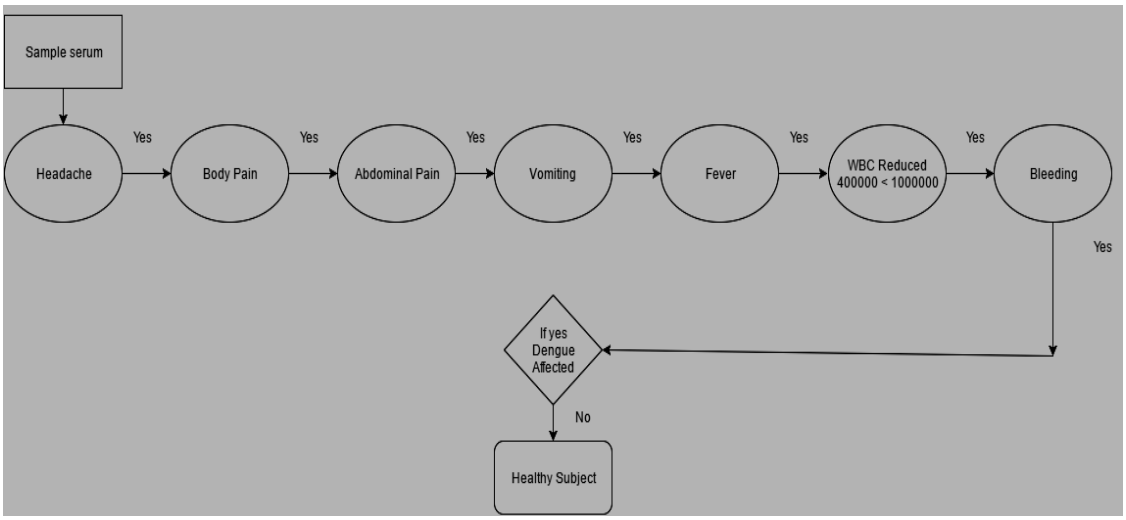


Fig. 2: Sequence of diagnosis performed on a test sample

Session is created and its connected to workflow by using Workflow designer in informatica. The Fig.3 below shows the session connected to workflow, this workflow is used to run the overall process for final output.

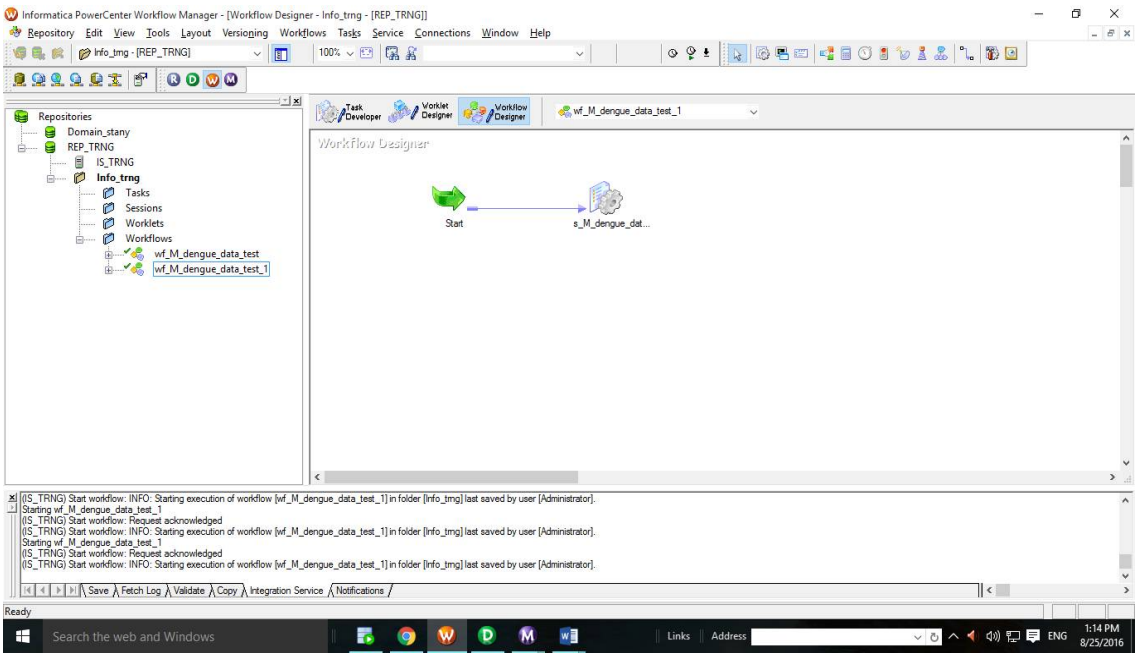


Fig.3: Workflow Designer

IV. RESULTS

70% of the 100 collected serum samples 70 % were found to be healthy subjects. Whereas the 30 % were found be affected with Dengue fever. As per our analysis below the shocking factor is Children below 10 years of age are prone to have dengue affected as compared to adults, hence precautionary measures have to be taken to control the number of death due to dengue. Analysis of collected data samples are shown in Fig. 4.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

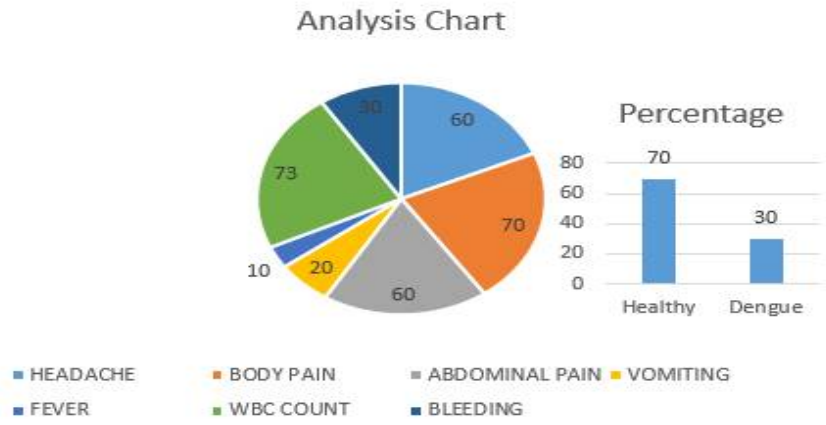


Fig. 4: Analysis chart of collected samples

Data are stored in oracle database. An Oracle database is a collection of data which can be called as a unit. Oracle database is used for two purposes: to store and to retrieve the data [13], it differentiated the dengue patient from healthy object for better view. Fig. 5 shows the patients who are all with positive results are painstaking to be affected on dengue disease and got segregated from healthy objects.

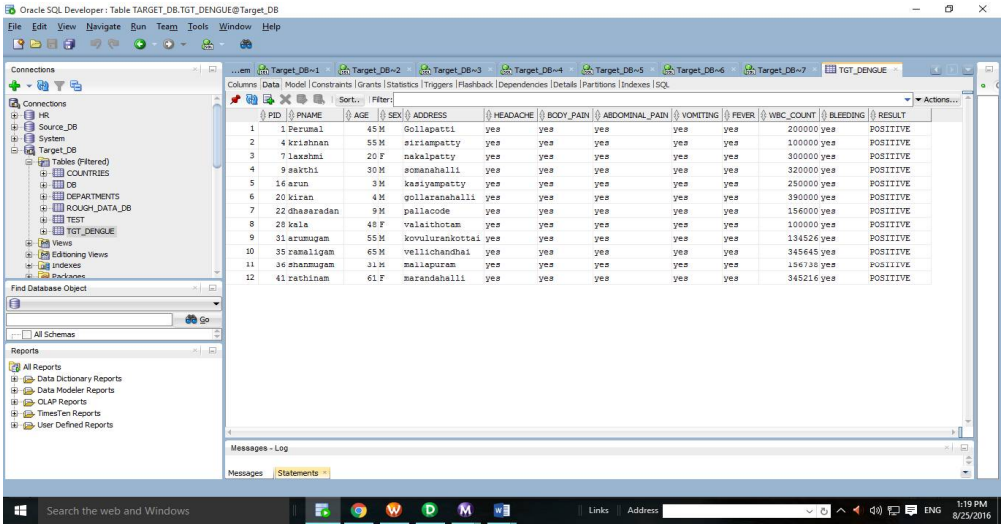


Fig.5: Patient positive report database

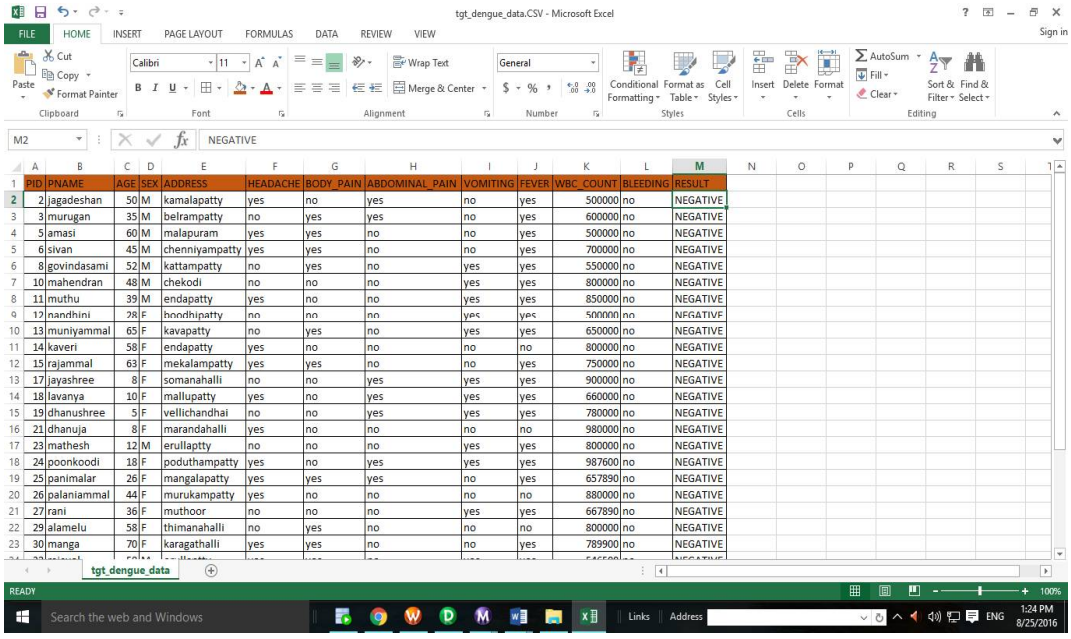
Fig.6 shows the patients with negative results are stored in flat file for future analysis. A flat file database is a collection of data which is stored locally on a host. The file is usually in an un-structured format. This evidence will help the doctors to view the particular person medical report from the first stage.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016



PID	PNAME	AGE	SEX	ADDRESS	HEADACHE	BODY_PAIN	ABDOMINAL_PAIN	VOMITING	FEVER	WBC_COUNT	BLEEDING	RESULT
2	jagadeshan	50	M	kamalapatty	yes	no	yes	no	yes	500000	no	NEGATIVE
3	murugan	35	M	belrampatty	no	yes	yes	no	yes	600000	no	NEGATIVE
5	amasi	60	M	malapuram	yes	yes	no	no	yes	500000	no	NEGATIVE
6	sivan	45	M	chenniyampatty	yes	yes	no	no	yes	700000	no	NEGATIVE
8	govindasami	52	M	kattampatty	no	yes	no	yes	yes	550000	no	NEGATIVE
10	mahendran	48	M	chekodi	no	no	no	yes	yes	800000	no	NEGATIVE
11	muthu	39	M	endapatty	yes	no	no	yes	yes	850000	no	NEGATIVE
12	nandhini	28	F	boodhipatty	no	no	no	yes	yes	500000	no	NEGATIVE
13	muniyammal	65	F	kavapatty	no	yes	no	yes	yes	650000	no	NEGATIVE
14	kaveri	58	F	endapatty	yes	no	no	no	no	800000	no	NEGATIVE
15	rajammal	63	F	mekalampatty	yes	yes	no	no	yes	750000	no	NEGATIVE
17	jayashree	8	F	somanahalli	no	no	yes	yes	yes	900000	no	NEGATIVE
18	lavanya	10	F	mallupatty	yes	no	yes	yes	yes	660000	no	NEGATIVE
19	dhanushree	5	F	vellichandhai	no	no	yes	yes	yes	780000	no	NEGATIVE
21	dhanuja	8	F	marandahalli	yes	no	no	no	no	980000	no	NEGATIVE
23	mathesh	12	M	erullaptty	no	no	no	yes	yes	800000	no	NEGATIVE
24	poonkoodi	18	F	poduthampatty	yes	no	yes	yes	yes	987600	no	NEGATIVE
25	panimalar	26	F	mangalapatty	yes	yes	yes	no	yes	657890	no	NEGATIVE
26	palaniammal	44	F	murukampatty	yes	no	no	no	no	880000	no	NEGATIVE
27	rani	36	F	muthoor	no	no	no	yes	yes	667890	no	NEGATIVE
29	alamelu	58	F	thimanahalli	no	yes	no	no	no	800000	no	NEGATIVE
30	manga	70	F	karagathalli	yes	yes	no	no	yes	789900	no	NEGATIVE

Fig. 6: Healthy objects are stored in flat file

V. CONCLUSION

Dengue is a life threatening diseases, widely prevalent in India since 1956. The dengue affected rate was 10,097 in 2014 with 37 deaths. In 2016, 41 lives have already been lost due to dengue and 19,704 cases were positive for dengue. The rapid diagnose kit was used for diagnosing dengue. However the results were unreliable.The Union Health Ministry has requested Indian Council of Medical Research to ban devices due to improper results. We are now left with no choice other than taking precautionary measures which involve making our environment clean and tidy which in turns annihilates the reproduction of Aedes mosquitos. Vaccination has to be developed for preventing this fatal disease and the same should be available for people at low cost.

REFERENCES

1. WHO SEARO Comprehensive Guidelines for Prevention and Control of Dengue and Dengue Haemorrhagic Fever Revised and expanded 2011.
2. Srinivas, Vaddadi, and Vaddadi Radha Srinivas. “DENGUE FEVER: A REVIEW ARTICLE”. INDIAN J MED RES (2011).
3. Mohan, D. K., and M. Dhananjaya. “A study of clinical profile of dengue fever in a tertiary care teaching hospital”. Sch J App Med Sci 1.4 (2013): 280-282.
4. Gunasekaran, P., et al. “Dengue disease status in Chennai (2006-2008): A Retrospective Analysis”. The Indian Journal of Medical Research 133.3 (2011): 322.
5. Fathima, A. Shameem, D. Shameem, and Nisar Hundewale Manimegalai. “A Review of Data Mining Classification Techniques Applied for Diagnosis and Prognosis of the Arbovirus-Dengue”. IJCSI International Journal of Computer Science 8.6 (2011).
6. Shaukat, Kamran, et al. “Dengue Fever Prediction: A Data Mining Problem”. Journal of Data Mining in Genomics & Proteomics 2015.
7. Faisal, Tarig, Fatimah Ibrahim, and Mohd Nasir Taib. “Analysis of significant factors for dengue infection prognosis using the self-organizing map”, 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2008.
8. Gupta, Nivedita, et al. “Dengue in India” The Indian Journal of Medical Research 136.3 (2012): 373.
9. Sharma et al., International Journal of Advanced Research in Computer Science and Software Engineering 4(6), June - 2014, pp. 1-6.
10. Malewar, Rahul. Learning Informatica Power Center 9.x. Packt Publishing Ltd, 2014.
11. Berson, Alex, and S. Smith. “Data Warehouse, Data Mining and OLAP”(1997).
12. Vadrevu, Keshav. “Informatica Power Center - The Complete Reference: The one-stop guide for all Informatica Developers” (2014).
13. Lorentz, Diana. “Oracle Database SQL Reference, 10g Release 2 (10.2) B14200-02”.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

BIOGRAPHY

Mrs. S.STANY LEENA PRINCY, Research Scholar, Department of Computer Science, Don Bosco College, Dharmapuri, Tamil Nadu, India. She has received her Bachelor's degree in computer science from Thiruvalluvar University, Vellore in 2010 and also holds her MCA degree from SRM University, Chennai in 2013. She worked as a System Configuration Specialist in Aon Hewitt, Chennai from (2013-2014). Her research area of interest includes Data mining and Database management system.

Mr. A. MURUGANANDAM, Assistant Professor, Department of Computer Science, Don Bosco College, Dharmapuri, Tamil Nadu, India. He is a Research Scholar in the field of Wireless Sensor Networks at Bharathiyar University, Coimbatore, Tamil Nadu, India. His research is focusing on Secure Data for preventive and selective jamming attacks in Wireless Sensor Network.