

# Disease Prediction using Enhanced Hybrid Algorithm with Manifold Dimensional Data

Dhivya S, Anguraju K, Suvitha K, Preethi P, Saravanabhavan C



**Abstract:** Medicinal services industry has become huge business. The medicinal services industry creates a lot of social insurance information day by day that can be utilized to separate data for foreseeing malady that can happen to a patient's future while utilizing the treatment history and wellbeing information. This concealed data in the human services information will be later utilized for full of feeling basic leadership for patient's wellbeing. Additionally, this region needs improvement by utilizing the educational information social insurance. Significant test is the manner by which to extricate the data from these information on the grounds that the sum is enormous so a few information mining and machine learning systems can be utilized. Additionally, the normal result  $a-n$  extent of this venture is that in the event that malady can be anticipated, at that point a treatment can be given to the patients which can lessen the danger of life and spare existence of patient and cost to get treatment of ailments can be decreased up somewhat by due acknowledgment. For this expert problem, a probabilistic displaying and profound learning approach will prepare a Long Short-Term Memory (LSTM) repetitive neural network and two convolutional neural systems for forecast of illness which is named as an upgraded half and half algorithm. The fast adoption of electronic wellbeing records has made an abundance of new information about patients, which is a goldmine for improving the understanding of human health. The above strategy is utilized to anticipate epidemic ailments utilizing quiet treatment history and wellbeing data. The parameter thought about is that this mode lessens test and train set blunder and increment information enlargement with better accuracy of 73.6%.

**Keywords :** Healthcare, LSTM, Epidemic, Inhabitants.

Manuscript received on February 10, 2020.

Revised Manuscript received on February 20, 2020.

Manuscript published on March 30, 2020.

\* Correspondence Author

**Dhivya S**, Assistant Professor, Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, Tami Nadu, India.

**Anguraju K**, Assistant Professor, Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, Tami Nadu, India.

**Suvitha K**, Assistant Professor, Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, Tami Nadu, India.

**Preethi P**, Assistant Professor, Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, Tami Nadu, India.

**Saravanabhavan C**, Head of the Department, Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, Tami Nadu, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## I. INTRODUCTION

**A.** Information mining in databases is a programmed extraction of understood and intriguing examples from huge information assortments. Information mining is a course at the coordination of measurements and software engineering and is the technique that endeavors to decide designs in the enormous datasets. It utilizes strategies at the blend of AI, man-made brainpower, database the executives frameworks and insights. The general target of the mining procedure is to pull out data from a dataset and convert it into an understandable structure for future use [1]. Information mining is a strategy for sharp through colossal measures of information for patterns. The significant goal of information mining is to pull out significant data from datasets which was not some time ago known. It is regularly used to perceive certain examples or trends. One critical factor of information mining is it will often be used to dissect information from verity of unique perspectives. The basic data that I accomplished from information mining might be used to support up benefits or lesser costs. Data mining is a sensible technique that is used to look through tremendous measures of information so as to discover huge data. The goal of this framework is to find designs that were not known previously. Once we have discovered these examples, we can use them to determine various issues.

## B. OBJECTIVES

This research work is based on some multi dimensional algorithms in data mining for prediction. After analyzing the advantages and disadvantages of different existing algorithms, a unique hybrid algorithm for disease prediction with real time data clusters is proposed. Five different problems were investigated and solved. Then a new hybrid algorithm for disease prediction was also developed [2].

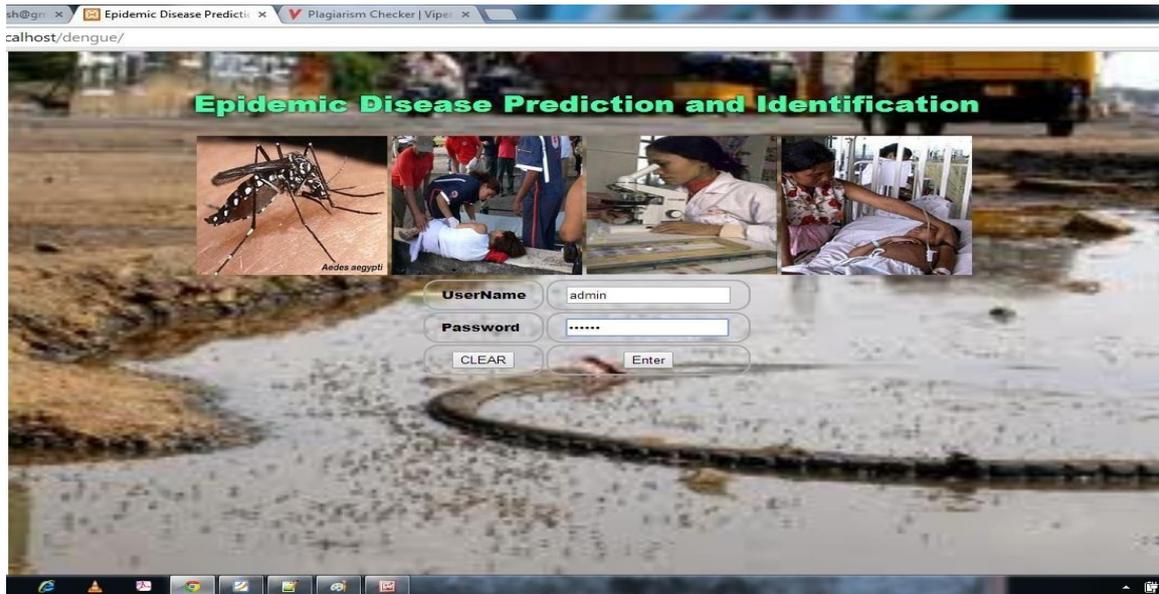
1. The main aim of this study is to identify the change of disease hit using the classification rule in data mining.
2. The disease prediction is carried out with the help of clustering algorithm.
3. This study made a mechanism for disease prediction using ID3 algorithm of decision tree.
4. The objective is to identify the correlating attributes and predict the disease using A prior concept in association rules.

5. A comparative study of different data mining algorithm used in prediction model was, analyzed and the advantage and limitations were studied.

A new hybrid algorithm was developed and its software implementation was done using PHP.

The major goal of the method is to generate an easy, firm and an effectual algorithm for disease forecast, with fewer fault rates then able to relate it with even large data types and display rational forms with contingent modifiers. A new hybrid algorithm was designed in data mining to identify and predict diseases. The Advanced Epidemic Disease Prediction and Identification (AEDPI) algorithm is the amalgamation of

classification, decision tree, association rules and clustering is used to forecast the probability getting disease in some specified areas. It also displays the relationships among different parameters for the disease prediction. This process was implemented by using a novel software program developed using PHP. Furthermore, comparison is made with a wide range of modern manifold learning algorithms, i.e. as Principal Component Analysis (PCA), Multidimensional Scaling (MDS), and isometric feature mapping (Isomap). Finally, we draw conclusion and discuss about future work in the last section.



**Figure 1.AEDPI-login Page**

### C. METHODOLOGY

Information mining can adopt on various strategies and fabricate various models relying on the kind of information included and the destinations. Various information mining calculations on multi dimensional information investigation utilized right now. Affiliation rules, Clustering strategies, Decision tree, Classification Rules and Statistical mining devices are utilized as regular models for expectation in information mining. In information mining, for the infection expectation and recognizable proof another mixture calculation was designed. The Epidemic Disease Prediction and Identification (AEDPI) calculation, is a blend of choice tree and affiliation rule mining to foresee the progressions of getting Epidemic Disease in some chose areas. The forecast of sickness right now be appeared by the connections between wanted parameters. The execution of the calculation is created in PHP language[3]. At the hypothetical level, we made a non-direct information digging model for connecting factors utilizing Decision tree calculation. By crossing the Decision tree from root to leaves the forecast rules is straightforwardly gotten. The sensible reliance between different qualities of an element utilizing affiliation rule of An earlier rule is developed. Affiliation qualities are estimated by estimating the certainty and backing. Utilizing a base certainty and bolster edges, the standard mining calculation

distinguishes all affiliation fulfilling the predefined parameters and discover the conditions between various traits of the equivalent entity. We can apply the guidelines and information to a measurable procedure in the group examination to extricate every single imaginable bunch from unlabelled data. The results can be spoken to in graphical structure for investigation. The Usage of this was done in 2 phases.

### D. PHASE 1:

Made a non-direct Data Mining model by utilizing Decision tree and built up a forecast rule for co-relating parameters. Let  $T$  be the preparation informational collection with class labels  $\{c_1, c_2, \dots, c_k\}$  and  $X$  is the non-class properties of  $T$ . Structure the quality rundown of  $X$  w.r.t  $T$  a disorted the quality rundown and utilizing significance investigation made property evacuation utilizing the threshold. We Measured uncertainty coefficient of an attribute  $X$  using the equation,  $UC(x) = \frac{\text{gain}(x, T)}{\text{info}(T)}$  and  $\text{Gain}(x, T) = \text{info}(T) - \text{fo}(x, T)$ .

**E. PHASE 2:**

Locate the incessant things and put away in the smaller structure. Blended the sets and enlisted as a tally if different transaction. Find all the successive things and their help by examined the database. A tree was made with a root hub as invalid. Expelled all non-visit items, and recorded the staying as per the request in arranged incessant things relying on the principal exchange from the database. Used the exchange to develop the primary part of the tree with every hub compares to visit item. Removed non-visit things and embedded in the tree, and expanded thing tally until all exchanges were finished relying on the following exchanges.

**F. TECHNICAL DETAILS OF THE ALGORITHM**

1. At the intangible level, by Decision Tree, create a non-linear data mining prototype for associating parameters. Mining the past data on family history, a prediction can make such that the high risk patient have a heart attack.
2. Develop the prediction rule, by traversing the tree from root to leaves.
3. Find the logical need between several attributes of an object using association rule of common path growing algorithm. In A prior algorithm, bottom up approach is used. If the frequent item set becomes longer, the algorithm has to go through many iterations. So the performance may decrease. In order to avoid this disadvantage, bi-directional search is incorporating here which takes advantages of both bottom up and top down process. For getting maximal frequent sets, in each pass, it counts the support of candidate in bottom up direction and top down approach. Subsequent to making the rules, find the intelligent reliance between different properties of a substance utilizing affiliation rule of Apriori principle. The quality of affiliation can be estimated by ascertaining the certainty and backing. Using the minimum sureness and support edge, the rule mining algorithm identifies all associations satisfying the specified limits. It is also cast off to find the needs between diverse attributes of the similar entity. The association rule for 'dengue hit' and 'no hit' using the attributes age group, area, sex, population and climate are expressed as  $Age(x, (24, 30)) \wedge area(x, i'urban') \wedge population(x, high) \wedge climate(x, past \text{ in } onsoon) [4\% \wedge 30\%]$
4. Applied the rubrics and data to arithmetical technique in cluster study to cutting all possible clusters from unlabelled data.
5. Used activity indicators to improve the readability. As more than one predicates can be involved here, there is no prefixed order for predicates.
6. The results of the analysis can be represented in graphical form.
7. This algorithm is a hierarchical or multi level hybrid algorithm which act as the combination of association, decision and clustering method.
8. Rare association rules and negative association rule is also considered while implementing this hybrid algorithm.

**G. COMPUTATIONAL COMPLEXITY**

Computational complexity depends on the number of attributes. Using utility measure 1-item transaction can be separated out and mined. If the number of input attributes is small, the time complexity is linear in the number of records. If the rule has Two or more predicates then it can called as high dimension to multiple predicates. To apply association rules we have to extract high frequency symptoms first, Which corresponds to 2 item sets. After eliminating cases using minimum support and minimum confidence, we can proceed to higher combination. Rare association rule also has to be considered for computational complexity. Absolute rarity can be determined by a cut off threshold limit. Relative rarity can be opposite of frequent association.

This new fusion algorithm, Widespread disease prediction and identification (EDPI) algorithm is considering negative association also.

**H. RESULTS AND IMPLEMENTATION**

**Algorithm implementation screen**

Figure 2 shows the entry screen of epidemic disease prediction and identification software. The user or admin can enter to the software by entering accurate password.

Figure 3 shows the tabular representation of disease and age. There are different menus in this software such as entry, profile, analysis, criteria etc. Figure 4 shows the graphical representation of disease-age and disease-place in the data analysis screen of the newly created EDP software.

This software also helps to analyze the relationship among different Attributes such as gender, area etc. with the disease. The graphical as well as tabular data analysis can be done through this software. The user has to enter his profile through a user friendly menu. After entering his details, an analysis report will be displayed or printed, which shows the prediction details showing if there is any chance to incident any of the selected disease to that particular person. The entire analysis report is also visible to the admin. By referring table-1 and figures 1-8 with use the AEDPI algorithm, we can predict that one from tribal area, male between 0-20 and 11-20 having poor or average sanitation, taken food from outside and open source of water have a chance to the disease Dengue. The algorithm also predicts that the attributes hereditary, income and nature of job have not so much influence for the prediction of dengue.



Figure 2.Enhanced Hybrid

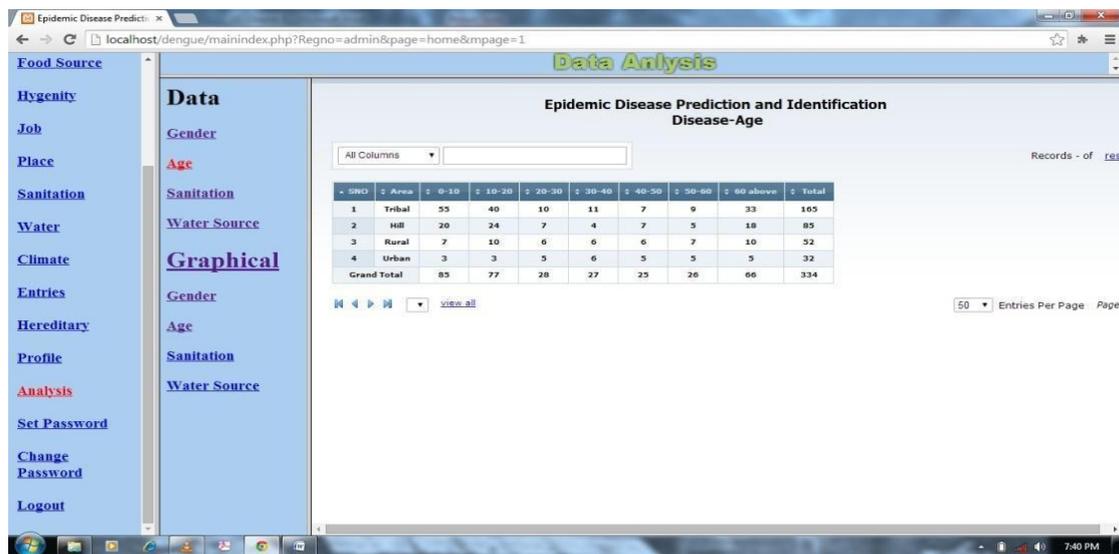


Figure 3. Disease-age tabular representation

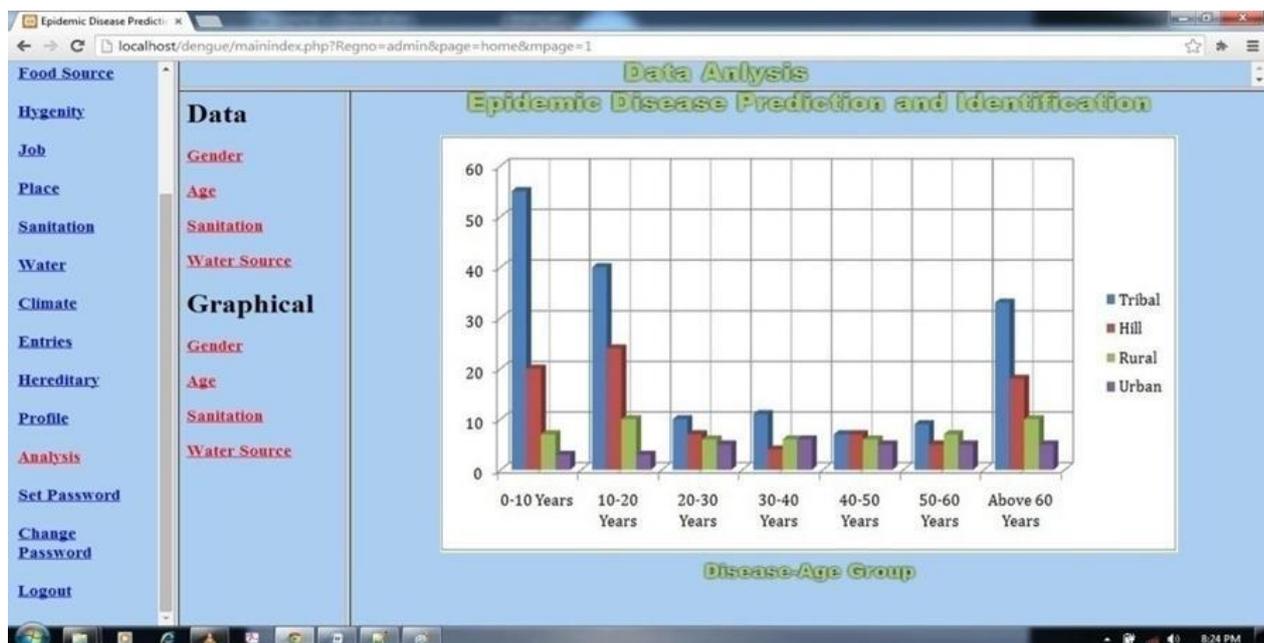


Figure 4.Disease-Age Group graphical representation

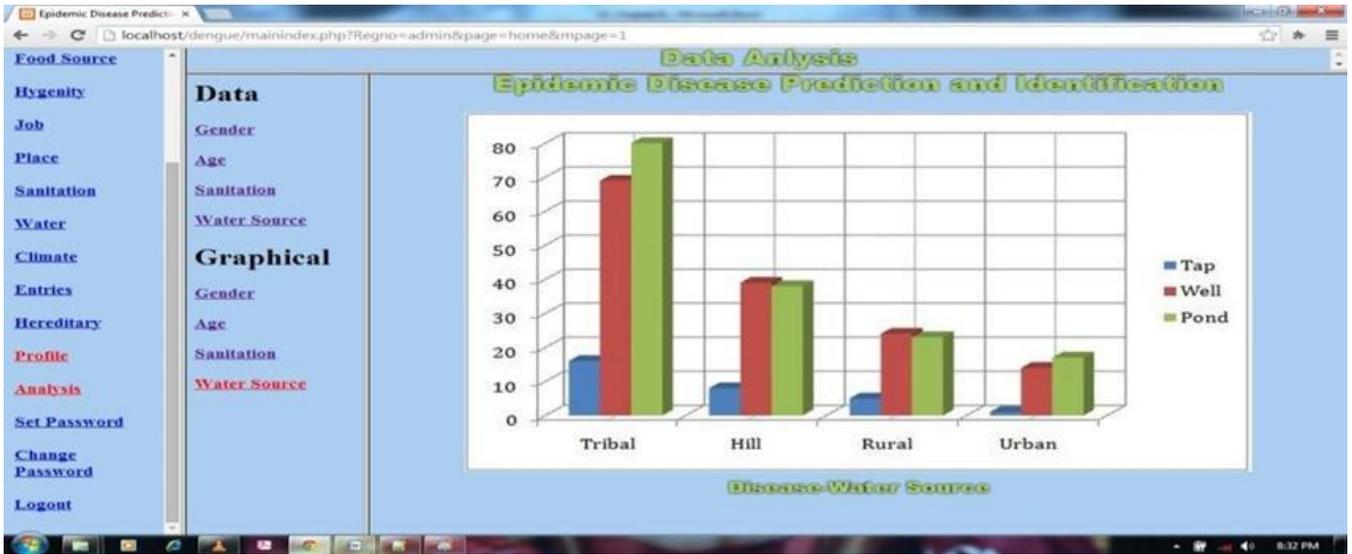
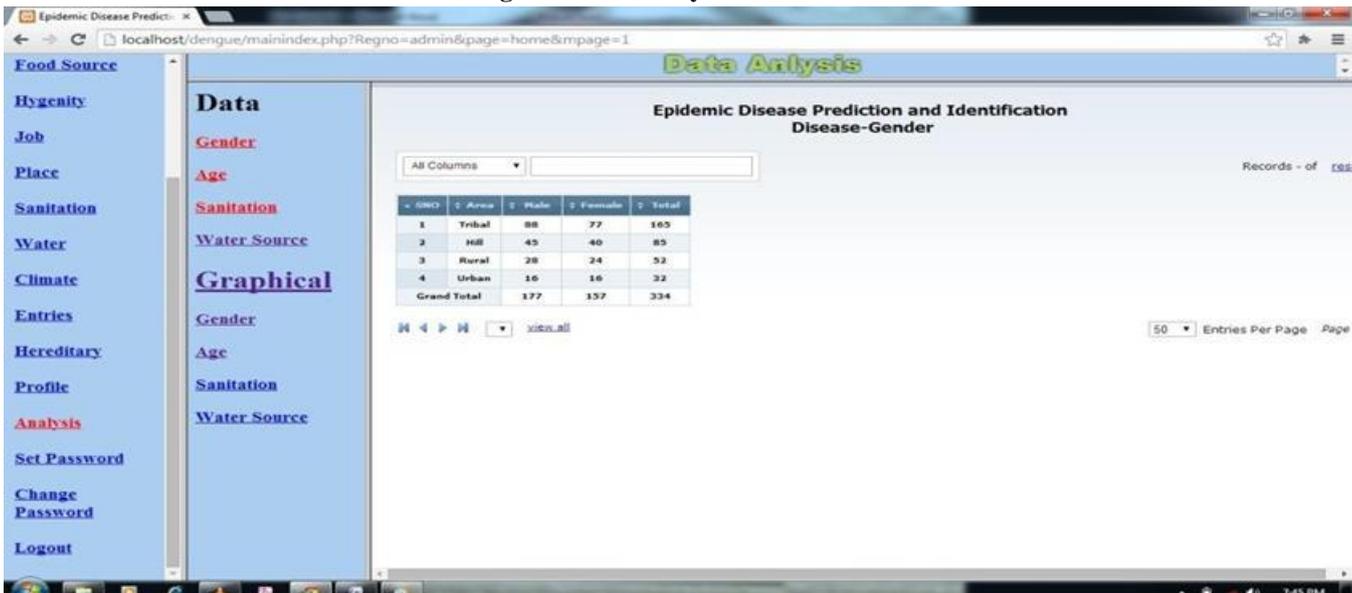


Figure 5. Graph output screen



Figure 6. Data analysis screen



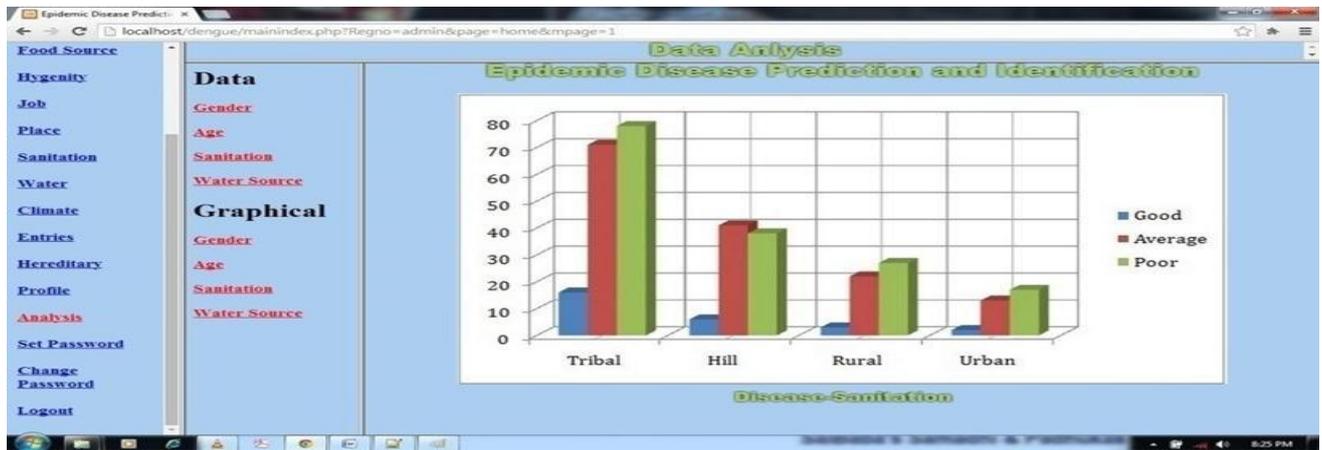


Figure 7. Disease-Gender tabular representation screen

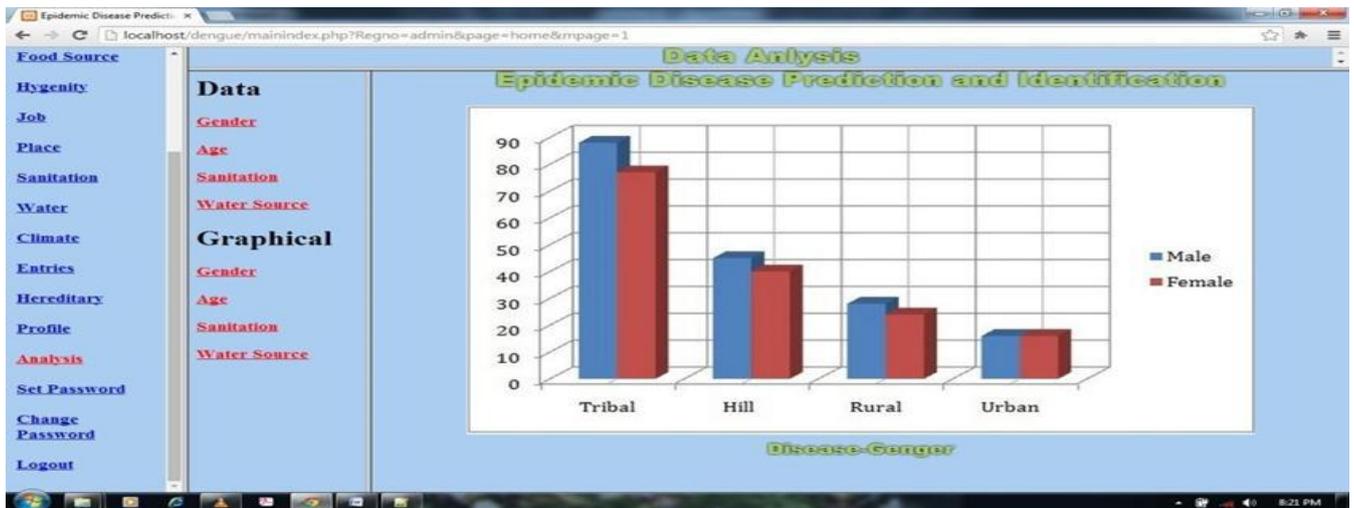


Figure 8 Disease-Gender Graphical representation screen

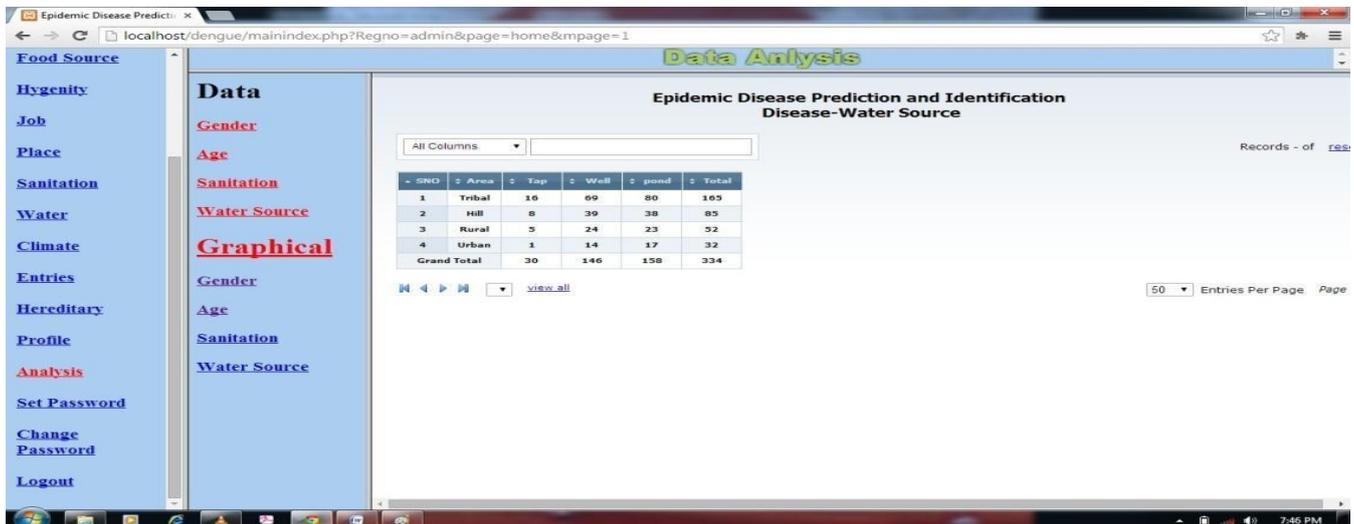


Figure 9. Disease-Water Source tabular representation screen

Similarly we can see that the chance for dengue fever for a person in the age group of above 60 years lives in rural area. By referring the nature of food habits and gender we can conclude that it affects more on male with a poor or average sanitation. Income, nature of job or family disease history and food habit has no impact here. We can also see that children in the age of 0-10 and female inhabitants in hill area, who taking food from outside having non-vegetarian food habits have more

tendencies to incident dengue. Also the attributes source of income and nature of job has no relation for the prediction of dengue. Study also shows that dengue is not a hereditary disease.

II. CONCLUSION

The planned new fusion algorithm is unique from the usually used prediction algorithms in data mining. The planned methods overawed the difficulties of existing methods as the number of frequent items is less. The novel algorithm proved to be efficient in terms of complexities such as time and space and proves to be exact when related with a standard arithmetical analysis tool such as SPSS the prediction can also reveal distinct group of patients having a common illness with various

treatment plans. It can also give common causative factors and other deficiencies.

The outliers as well as clusters can also be detected graphically using this algorithm and all extreme outliers can be further analyzed using root cause analysis if an assignable cause can be found. The algorithm has been implemented successfully by developing software using the programming language PHP. The results are tested and verified. The proposed new method is unique and different from the ancient and traditional combined methods.

Table-1 results obtained from EDPI

Disease	Area	Gender	Age Group	Sanitation	Food Source	Water Source	Type of food	Hereditary
Dengue	Tribal	Male	0-10,10-20	Avg, Poor	Outside	Pond	Non-veg	No
Dengue	Hill	Female	10-20,60Above	Avg, Poor	Outside	Pond	Veg	Yes
Dengue	Rural	Male	0-10,10-20	Poor	Home	Well	Non-veg	No
Dengue	Urban	Female	0-10,i60iAbove	Poor	Outside	Well	Veg	Yes

REFERENCES

1. S. Chaudhuri, S. M. Nelson, M. Goldbaum Detection of blood vessels in retinal images using two-dimensional matched filters IEEE Transactions Medical Imaging, 8(1989), pp. 263-269.
2. Goldbaum, M., Taylor, A., Chatterjee, S., Boyd, J., Hunter, E., and Jain, R. (1996). Automated diagnosis and image understanding with object extraction, object classification, and inference in retinal images. Proceedings of 3rd IEEE International Conference on Image Processing 3, 695-698.
3. V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al.
4. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs JAMA, 316(2016), pp. 2402-2410
5. C. S. Lee, D. M. Baughman, A. Y. Lee Deep Learning Is Effective for the Classification of OCT Images of Normal versus Age-Related Macular Degeneration Ophthalmol. Retina, 1(2016), pp. 322-327
6. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In 2016 IWWW Conference on Computer Vision and Pattern Recognition, pp. 2818-2826.

Trichy. She has 4 years of teaching experience. Her area of interest lies in Cloud Computing, Software Testing and published 4 papers (Scopus: 1) in international journals and presented 8 papers in national and international conferences in that area.



**Ms. P. Preethi**, Assistant Professor in the Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, Tamil Nadu, India. She received B.Tech., degree from Roever Engineering College, Perambalur in 2012. She was awarded with M.E., from Srinivasan Engineering College, Perambalur in 2014. She has 6 years of teaching experience and pursuing Ph.D., as part-time research scholar in Anna University, Chennai. Her area of interest lies in Cloud Computing, Network Security and published 11 papers (Annexure 1: 2 papers) in international journals and presented 13 papers in national and international conferences in that area.



**Dr. C. Saravanabhavan**, Head of the Department in the Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, Tamil Nadu, India. He received M.Tech., from Sathyabama University, Chennai in 2007. He has 15 years of teaching experience and completed Ph.D., in Anna University, Chennai. His area of interest lies in Data Mining, Cloud Computing, Network Security and published 21 papers in international journals and presented 23 papers in national and international conferences in that area.

AUTHORS PROFILE



**Mr. K. Anguraju**, Assistant Professor in the Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, Tamil Nadu, India. He received M.E., from Vinayaka Mission KV Engineering College, Salem in 2013. He has 6 years of teaching experience. His area of interest lies in Networking, Data Mining, Cloud Computing and published 6 papers in international journals and presented 5 papers in national and international conferences in that area.



**Ms. S. Dhivya**, Assistant Professor in the Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, Tamil Nadu, India. She received B.E., degree from Kongunadu College of Engineering and Technology, Trichy. She was awarded with M.E., from Kongunadu College of Engineering and Technology, Trichy. She has 4 years of teaching experience. Her area of interest lies in Cloud Computing, Network Security and published 4 papers (Scopus: 1) in international journals and presented 9 papers in national and international conferences in that area.



**Mrs. K. Suvitha**, Assistant Professor in the Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, Tamil Nadu, India. She received B.E., degree from M. Kumarasamy College of Engineering and Technology, Karur. She was awarded with M.E., from Kongunadu College of Engineering and Technology,