International Journal of Scientific Research in Science, Engineering and Technology (www.ijsrset.com)
© 2020 IJSRSET | Volume 7 | Issue 3 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
DOI : https://doi.org/10.32628/IJSRSET207365

Ontology Based System for Prediction of Diseases

Pallavi Laxmikant Chavan, Prof. Mandar S. Karyakarte
Vishwakarma Institute of Information Technology, Department of Computer science and Engineering, Kondhwa (BK), Pune, Maharashtra, India

ABSTRACT

In today's world of lifestyle, biomedical and healthcare act the main role through which disease in the patient can be identified. However, the current solution focuses on communities where the accurate prediction plays a major role to find out risk of the disease in the patient. The detection of disease is done by using prediction algorithm. Here, machine-learning algorithm is has been used to find the accuracy. The dataset has been is collected from certain hospitals and pre-processed where the missing values have been reconstructed before prediction process. Due to the huge amount of information in healthcare, the accurate result is the need for disease recognition and services. Generally raw data has bad quality because it does have exactness, completeness of records fields. Moreover, there would be different exhibits in different regions, the appearances of certain diseases, which may also weaken the prediction of the disease outbreak. Using the health record, our system received the rate of accuracy is 97%. In this proposed system, we provides prediction of various diseases that occurs through using machine learning that will be effective. In urban lifestyle, modern large cities have significant adverse effects on health, & increasing risk of diseases.

Keywords: Disease, Management Problem, Healthcare Data, Accuracy

I. INTRODUCTION

1. Challenges of current disease management programs

- Today, the world has become a challenge due to unhealthy lifestyles which result in aging problems. On the report of healthcare, millions of people die due to some disease problems, as a result, there is a need for effective strategies that are essential to prevent and manage. There is 72 program carryout for chronic disease management [4].
- Prevention and management of ill patients required a long period to note the patient's condition, the risk would be, and treatment required. The action to improve the management which may contain monitoring, checking the progress, environmental change which improves the health, the services, and the community programs. However, the current program focus on highly specialized cases. In order to improve the continuity, efficiency, and quality of the health care services integrated program are required.

2. The role in the model

- The key principle of the management program is to make the process easy and short term. The ontology prediction system helps out the people to be aware of if they carryout same lifestyle then they will be a part of the disease. Even if they are not part of it now and in case if they will not change their lifestyle accordingly then they will be a part of disease sooner or later if they don't change their lifestyle. During this process, the patient lifestyle is reviewed based on the features

that are collected, analyzed, and evaluated. There are lots of computerized tools which are lack in inaccuracy and inefficiency of the decision-making process.

3. The objective of this study

- To contribute to the personal care, and decision support system. The main aim is to develop a knowledge-based system for the patient, an ontology-based framework for prediction of the diseases. Prediction of the disease has been developed that is based on the patient lifestyle and the medical domain. Evaluation of technical validation and the functional usability of real-life patient analysis.

As stated by the medical report, the death rate of humans is high due to disease and around 70% of the money is spent over disease treatment. The disease is the main reason for death as per the report. Thus, the analysis of disease is more important to reduce such a problem. At present, there is a growth in technology which is paying some attention for disease prediction using analytics, where accuracy in the threat of disease is improved. The handling and processing of the measurements can been done, well manner by data analysis. The daily lifestyle varies day to day which may also result in a certain effect of the disease.

To solve the problem with both structured and unstructured format we determine the threat over disease in fields in healthcare.

First, we pre-processed collected data to find out the missing values and the missing values has been reconstructed that would help the prediction to define with high accuracy rate. Finally, we have used multiple algorithms to explore the prediction accuracy rate of the risk by disease. The models are processed with the combination of both structured and unstructured data from hospital records to see the predicted result. India is a developing country and population wise it is the second- largest country. The health care needs of its vast population and ensuring accessibility, efficiency, equity, and quality of healthcare as India faces a challenge to meet its needs [2]. The instrument is found useful for investigating patterns and determinants of health- promoting lifestyle, as well as the effects of interventions to alter lifestyle [3].There is a big challenge stands in front of health care organizations (hospitals and medical centers) is to provide quality service at affordable costs.

However, it is not that easy to describe static standards that meet all patient needs in such volatile environments. Here, decisions and management of status are usually dependent on the question and answering of the medical staff who are involved in the healthcare system. For large volumes of data, ontology is one of the proposed solutions to deal with it [9].

- Ontology is defined as shared knowledge, incorporates the relevant domain concepts and their associated relations. Here, the health status of the patient is identified by the ontology-based approach which is used by the work proposed a smart body environment framework.
- The misdiagnosis cases of the patient and allow dynamic reconfiguration of health disorders orders, it is important to develop a comprehensive representation of knowledge to capture the real context of the patient.

II. RELATED WORK

The domain uses the formal modeling structures by explicitly defining the commonly approved conceptual model like classes representing concepts, properties which describe attributes and relationships between the concepts, restrictions which specifies logical constraints on concepts like ontology facilitates knowledge sharing and reuse in an intuitive and machine-understandable way. Instantiated domain ontology is also known as a knowledge base, which together is problem-solving methods

constitutes the foundation of many intelligent applications and systems in knowledge-intensive domains. However, it has not still been exactly standard in this area as the structure of the knowledge is a heterogeneity modeled.

In addition, most of these ontologies involve complex modeling primitives, making it difficult for knowledge engineers as well as healthcare staff with no advanced programming skills to understand and manipulate. The development of hybrid guidelines of a central repository and development of a graphical framework for collaborative specification and maintenance of guideline- based knowledge [5].

Here in this study, we have built the ontologies based framework on how it can be integrated and used. Support for the decision-making [6] process was a proposed method for representing medical knowledge. Multiple sources of patient data are demanded collected and analyzed to provide continuous and personalized disease management. A simple decision-making model based on patient specified data obtained from the health records.

III. MATERIALS AND METHODS

The model we design goes through different stages as shown in figure 3.1 of the stage:

PreParation:-

The Preparation is the Processing of the raw data and knowledge which would be used for the processing purpose. The patient data may come from multiple sources, hence normalization is required to reduce the heterography of the data. The record contains parameters like the physical work, food intake, lifestyle through which the result conceivable critical or not.

Pre-processing:-

Here, we have performed analyzing, grouping, removing encoding of data. The data that is collected from the hospital is the raw data to make for use it should be processed. Data is the group as per the need next is to remove unwanted data null type of data. For the prediction, the data is required to be encoded.

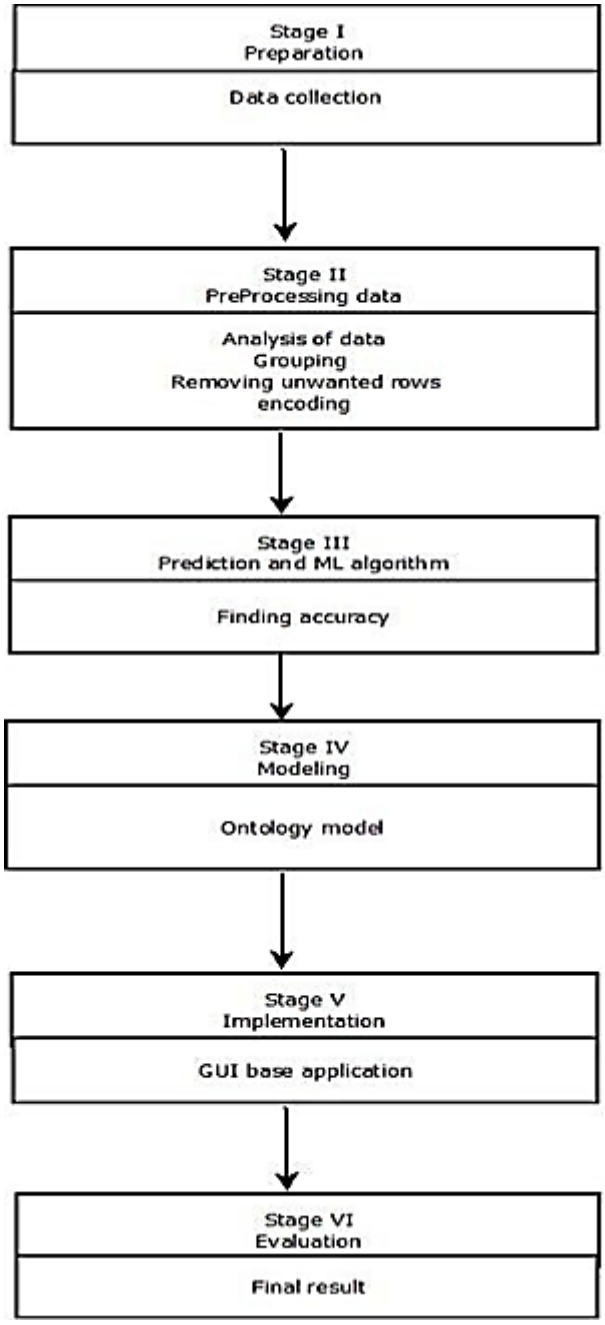

Fig. 3.1 Stage Diagram

Prediction and ML algorithm

The encoded data is used for the prediction of the data to find out how fast it works which gives high accuracy. The algorithm which takes less time and gives high accuracy is performing best on your data.

Modeling

The ontology base model is been developed for the prediction of lifestyle data. Ontology is the question answering base where it is the relationship-based system. *Implementation*

For the implementation of the system, we have used Django which response to the use. The Gui is made through the URL base it is accessed.

Evaluation

The final result of the prediction algorithm is displayed where there is the a critical patient or not. The data of the new patient is taken in accord with previous data whether the patient has the chance to gate the disease or not is predicted shown in figure 3.2.

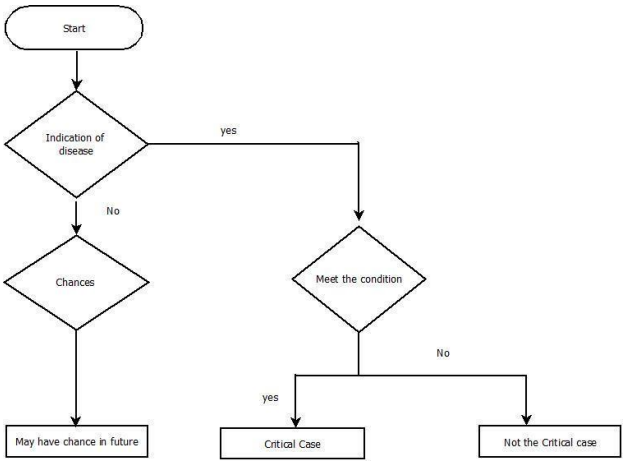

Fig. 3.2 Activity Diagram

IV. SYSTEM IMPLEMENTATION

To integrate data and achieve accessibility with the user application platform, we deployed an architecture figure to describe the architecture figure 4.1 of our system in three layers.

- Knowledge layer, which is the compression of the records, knowledgebase, domain base of the patient.
- Service Layer provides services like normalization, management, decision-making of the services. The decision support service, which is the compression of the knowledge layer helps for decision-making.
- Application Layer, support user application for proving of the standard services which invoke the services, retrieves required data of records and gave appropriate result. Currently, we have developed a URL based application.

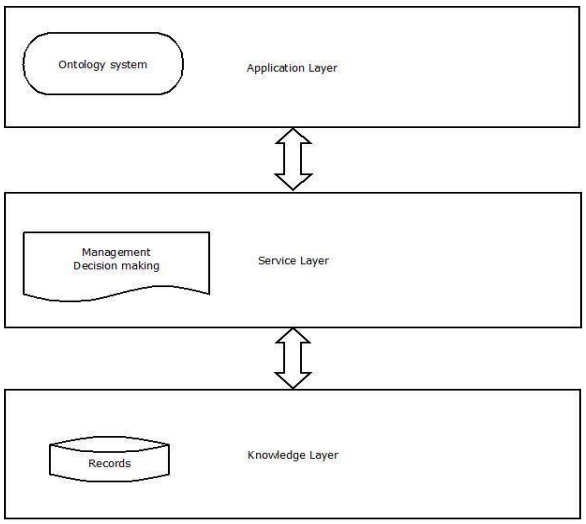

Fig. 4.1 Layer Diagram

V. SYSTEM EVALUATION

The combination of different factors that describe patient condition what has to happen or is. The several discrete values as the number of combinations should be taken into consideration. Data quality is one of the most important problems in data management since dirty data oftentimes leads to inaccurate data analytics results and wrong business decisions Poor data across businesses and the government cost. Not all possible combinations reduce significantly the total number of potential features that describe each patient [8]. We extract information about the disease type over their lifestyle and demographics including age, gender, and race. Next, we have followed the data organization and pre-processing steps. All past patient records are organized as follows.

- Reducing the Number of Features: The Features which are not of any use or not used for the predication should be removed to reduce the

population.

- Removing Patients With no Record: The patient who has no record should be removed since there is nothing on which a prediction can be based.
- Splitting the Data Into a Training Set and a Test Set Randomly: we have taken the 30% data as the test data and 70% data as the training data.
- Apply the algorithm: K-Nearest Neighbour, Random forest, Decision tree, support vector machine, and linear regression.
- Validation: Through use the algorithm the data is predicted to get an accurate result.

We have used the formula to find the accuracy of the data they are as follow:

True-Positive Rate = TP / TP + FN (1)

False-Positive Rate = FP / FP + TN (2)

True-Negative Rate = TN / TN + FP (3)

False-Negative Rate = FN / FN + TP (4)

Accuracy = TP/ (TP+TN)*100% (5)

Table: Symbol table

Symbols	Description
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

VI. APPROACH / ALGORITHM

In biodiversity, the compilation and cleaning of data

Decision Tree pseudocode:

1. Begin the tree with the root node, says S, which contains the complete dataset.
2. Find the best attribute in the dataset using Attribute Selection Measure (ASM).
3. Divide the S into subsets that contains possible values for the best attributes.
4. Generate the decision tree node, which contains the best attribute.
5. Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this needed for analyses and prediction of species distributions is a time-consuming process requiring a solid understanding of data formats and service APIs provided by biodiversity informatics infrastructures[9].

Begin with a simple, intuitive problem discussion of how ideas from Computational Learning Theory and On-Line Algorithms might be fruitfully combined and the task each day of predicting whether it will rain that day[10].

A survey of recent advances in the application of genetic algorithms to problems in machine learning. Genetic algorithms have also been applied to many traditional machine learning problems, including concept learning from examples, learning weights for neural nets, and learning rules for sequential decision problems [11].

Here, we have used a Decision Tree algorithm that gives good accuracy as compared to other algorithms. The algorithm used is the supervised type of algorithm i.e. it makes machine learn explicitly, data with clearly defined output is given, Predicts outcome/ future, resolves classification, and regression problems. Figure 6.1 process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

VII. RESULT AND DISCUSSIONS

The information about patients' healths through which the condition of the patient can be seen. This system obtains a step ahead of care and prevent medical mistakes also it saves their own lives at times such that it will reduce the diseases in society. Ultimately, it helps us to save the lives of people, improve care, and save a significant amount of money. Detailed of patient information will be taken and which would be evaluated whether it will be in a critical state or not as will satisfy it or not. If it is in critical state then the patient needs to change his/her lifestyle. The Prediction and analysis had been done on the data taken from the hospital. Decision tree algorithm has been used and it has given 97% good as compared to others like K- Nearest Neighbour- 90-96%, Random Forest-96%, support vector machine-67%, and linear regression- 71%.

The problem-solving method for the patient which is formalized as an ontology-based model with computer- interpretable and human-readable, include lifestyle result in prediction. Design to asses the patient with different healthcare conditions to model with specific criteria. The data classification is done on the comparison of critical condition or not, our evaluation scheme provides detailed and objective on the patient's condition. We have developed a service-oriented URL based model which widely acknowledged in healthcare, to realize the accessibility and utility of this model. This model can be used for decision-making and productive interaction between patients and healthcare providers. The framework provides the basic APIs and knowledge within the ontology system. We are working to solve this problem, to evaluate the accuracy of our system and the classification of records.

VIII. RESULTS OF ANALYSED DATA

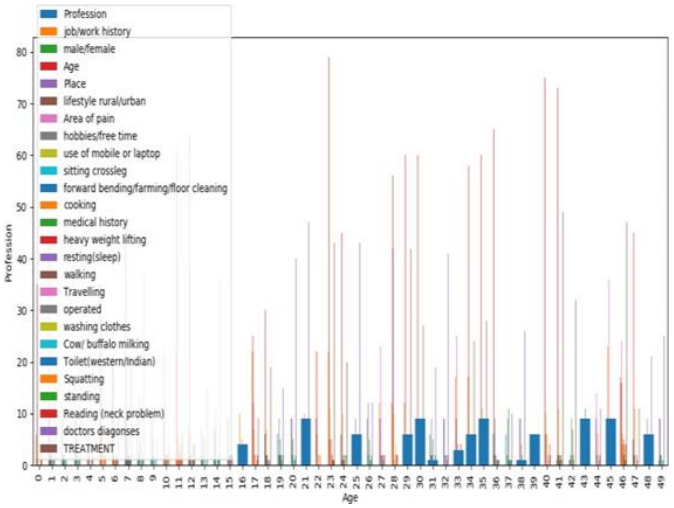

Fig. 7.1 Rawdata collected

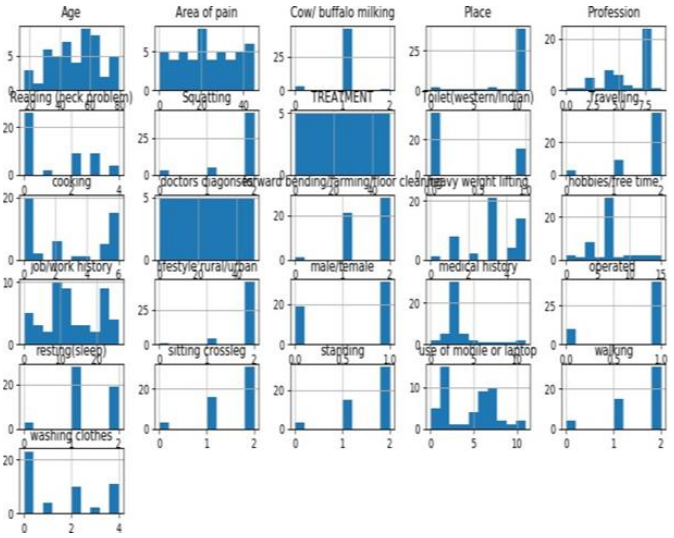

Fig. 7.2 Dataset histogram view

The raw data figure 7.1 is collected from the healthcare which is required to be processed. Figure 7.2 are the parameter in the data which to be described in the histogram format. The figure 7.3 shows the encoded data which is used for the farther processing purpose.

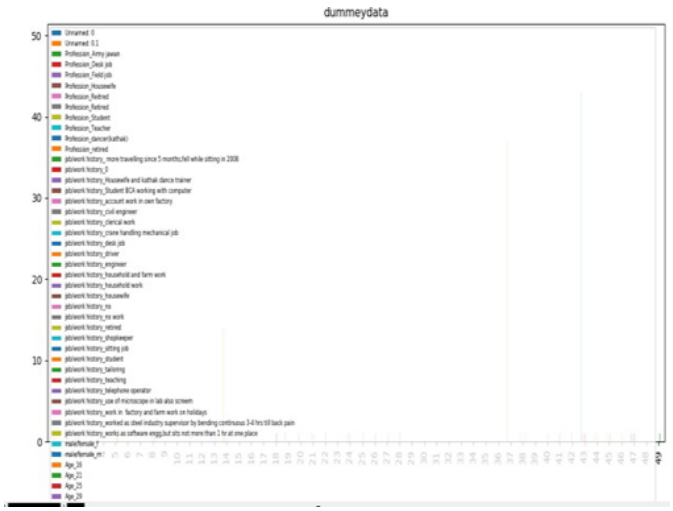

Fig. 7.3 Encode data representation

IX. CONCLUSION AND FUTURE WORK

Here, we have done how techniques from machine learning can be used for modeling prediction of disease. With the growing availability of data, the research focuses on machine-learning approaches that will continue to find accurate results. The algorithm reaches 97% with a decision tree algorithm as compared to several prediction algorithms to reach our best. In this domain of the healthcare and the medical, evaluation of the performance with reference to classification accuracy and interpretability are proposed. For knowledge injection into input features, objective functions, and output labels, we have implemented a different algorithms.

Therefore, the analysis over the chronic disease is more important to reduce the risk of people's life. The result obtained from the predictive algorithms like K-Nearest Neighbour- 90-96%, Random forest -96%, support vector machine- 67%, and linear regression-71%. Where Decision tree provides us with the highest rate of accuracy is 97% since they follow the random forest.

In this paper, we have proposed an ontology-based approach whereas technical level evaluation demonstrates validation of the model and functional level evaluation shows high accuracy and completeness of the decision made. The system will help in handling the tracked record of patients. The system changes the clinical practice which is the creative process to automatically test through ontology design.

As future work, we will use a proposed algorithm of machine learning and neural network algorithms like Naïve Bayes Classification, Ordinary Least Square Regression, Ensemble Methods, Apriori Algorithm, Principal Component Analysis, Singular Value Decomposition, Reinforcement or Semi-Supervised Machine Learning, Independent Component Analysis, RRNA,CNNA, etc.

X. REFERENCES

[1]. Ornish, D., Brown, S.E., Billings, J.H., Scherwitz, L.W., Armstrong, W.T., Ports, T.A., McLanahan, S.M., Kirkeeide, R.L., Gould, K.L. and Brand, R.J., 1990. Can lifestyle changes reverse coronary heart disease?: The Lifestyle Heart Trial. The Lancet, 336(8708), pp.129-133.

[2]. Roger Walsh. Lifestyle and mental health. American Psychologist, 66(7):579, 2011.

[3]. Susan N Walker, Karen R Sechrist, and Nola J Pender. The health-promoting lifestyle pro le: development and psychometric characteristics. Nursing research, 1987.

[4]. Edward H Wagner, Connie Davis, Judith Schaefer, and Michael Von Ko-r . A survey of leading chronic disease management programs: Are they. Managed care quarterly, 7(3):56{66, 1999.

[5]. Susan Gauch, Jason Cha ee, and Alexander Pretschner. Ontology-based personalized search and browsing. Web Intelligence and Agent Systems: An international Journal, 1(3, 4):219{234, 2003

[6]. Shusaku Tsumoto. A decison support system based on congenital mal-formation image

databases. In Medical Image Databases, pages 375{391. Springer, 1998.

[7]. C Warren Olanow, Ray L Watts, and William C Koller. An algorithm (de-cision tree) for the management of parkinson's disease (2001):: Treatment guidelines. Neurology, 56(suppl 5):S1{S88, 2001

[8]. Ihab F Ilyas, Xu Chu, et al. Trends in cleaning relational data: Consistency and deduplication. Foundations and Trends R in Databases, 5(4):281{393, 2015

[9]. Cherian Mathew, Anton Guntsch, Matthias Obst, Saverio Vicario, Robert Haines, Alan R Williams, Yde de Jong, and Carole Goble. A semi-automated work ow for biodiversity data retrieval, cleaning, and quality control. Biodiversity data journal, (2), 2014

[10]. Avrim Blum. On-line algorithms in machine learning. In Online algorithms, pages 306{325. Springer, 1998

[11]. John J Grefenstette. Genetic algorithms and machine learning. In Pro-ceedings of the sixth annual conference on Computational learning theory, pages 3{4, 1993.

[12]. David E Goldberg and John Henry Holland. Genetic algorithms and ma-chine learning. 1988.

[13]. Adele Cutler, D Richard Cutler, and John R Stevens. Random forests. In Ensemble machine learning, pages 157{175. Springer, 2012.

[14]. GASrard Biau, Luc Devroye, and GAAbor Lugosi. Consistency of ran-dom forests and other averaging classi ers. Journal of Machine Learning Research, 9(Sep):2015{2033, 2008.

[15]. Angela Neumeyer-Gromen, Thomas Lampert, Klaus Stark, and Gerd Kallischnigg. Disease management programs for depression: a systematic review and meta-analysis of randomized controlled trials. Medical care, pages 1211{1221, 2004

[16]. Clement Pimouguet, Melanie Le Go , Rodolphe Thiebaut, Jean Francois Dartigues, and Catherine Helmer. E ectiveness of disease-management programs for improving diabetes care: a meta-analysis. Cmaj, 183(2):E115{ E127, 2011

[17]. Michelle Longmire. Method, process and system for disease management using machine learning process and electronic media, October 1 2013. US Patent 8,548,828

[18]. Dragana Miljkovic, Darko Aleksovski, Vid Podpecan, Nada Lavrac, Bernd Malle, and Andreas Holzinger. Machine learning and data mining methods for managing parkinson's disease. In Machine Learning for Health Infor-matics, pages 209{220. Springer, 2016

Authors' Profiles

Pallavi Laxmikant Chavan received the Bachelor of Engineering degree in Computer Science and Engineering, from Shivaji University. She has completed M. Tech. in Computer Science and Engineering from Vishwakarma Institute of Information Technology, Kondhwa(BK), Pune, MS, India. Her area of interest is Data Science.

Mandar S. Karyakarte is fulltime professor for Computer Science at Vishwakarma institute of information technology, pune. He has been there from 2008. He had Lecturer, Bharati Vidyapeeth's College of Engineering for Women, Computer Engineering Department, 29/12/2003 to 23/06/2008.

Senior Lecturer, Vishwakarma Institute of Information Technology, Department of Information Technology since 24/06/2008 to 30/06/2009. Assistant Professor, Vishwakarma Institute of Information Technology, Department of Information Technology since 01/07/2009 to 29/03/2017 And Professor in Computer Engineering, Vishwakarma Institute of

Information Technology, Department of Information Technology.

Cite this article as :

Pallavi Laxmikant Chavan, Prof. Mandar S. Karyakarte, "Ontology Based System for Prediction of Diseases", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 7 Issue 3, pp. 277-285, May-June 2020. Available at
doi : https://doi.org/10.32628/IJSRSET207365
Journal URL : http://ijsrset.com/IJSRSET207365