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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Survey on On-line Medical Guidance

Madhuri N. Suryatal, Prof. P. S. Vidap

M.E Student, Dept. of Computer Engineering, Pune Institute of Computer Technology, SPPU, Pune, India

Assistant Professor, Dept. of Computer Engineering, Pune Institute of Computer Technology, SPPU, Pune, India

ABSTRACT: Health is important no matter what country we are from. Some important details like the patients allergy and medical history is overlooked by the doctors when they diagnosing. The procedure for diagnosing also consume a big amount of time although the patients comes with a minor problem. Patient can settle it by themselves if they have the right help. So, the technology can help by providing guidance to each patient. Instead of visiting hospital all the time if right help is provided to patients then they will save time and money. Each Internet user can do self-checking of his/her health based on certain symptoms with his/her queries. Then technology can pre-diagnose possible diseases and provide a risk warning and prevention.

KEYWORDS: Machine learning, CNN, RNN, Bayesian Classifier, Artificial Bee Colony, Naive Bayes and Jelinek-mercer smoothing.

I. INTRODUCTION

Health is a very important factor in human life so for protecting the health from various diseases humans are taking various kinds of help from the Internet. When humans do not feel well, nearly ninety percent of them first cross onto the Internet to search for associated scientific records. The Internet has already modified the Eco-device of clinical services for all important steps, inclusive of scientific session, clinic visits, remedy and recovery in addition to buying medication on-line.

Human Computer interaction is very powerful and important area of research because the world is moving towards the digitalized area. Thus there is a need for the digital system to correctly predict the diseases from patient's symptom. The population of world is increasingly day by day and there are very few doctors for every thousands of people.

The aim is to provide human-like, comprehensive and informative medical consultation. Detecting the diseases of a person by analyzing a symptoms and questions asked by patients is challenging problem because most people when they become sick, as a result of their lack of medical knowledge and experience, will describe their symptoms inaccurately in medical terms [1].

Deep learning based neural network models have achieved great success in many NLP tasks, including learning distributed word, sentence and document representation, parsing, statistical machine translation, sentiment classification[6], etc. Learning distributed sentence representation through neural network models requires little external domain knowledge and can reach satisfactory results in related tasks like sentiment classification, text categorization[13].

At present, the intelligent medical guide system has two main ways: 1) intelligent medical guide system which based on knowledge 2) intelligent medical guide system which based on similarity calculation[4]. Intelligent medical guide system which based on knowledge base, constructs the knowledge base by knowledge and experience of medical experts and uses fixed diagnosis condition to judge the disease which user may suffer from and to provide medical guide service in the form of expert system. Intelligent medical guide system which based on similarity calculation develops vector space model by symptoms which user input and symptoms of disease, judges the disease which user may suffer from and provides medical guide service using the similarity calculation [4].

II. REVIEW OF LITERATURE

The basic of online medical guidance depends on machine diagnosis so in 1991, R. S. Ledley and L. B. Lusted proposed idea of machine diagnosis. It generate list of the possible disease complexes that the patient can have that are

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

consistent with medical knowledge and the patients symptoms[3]. In 2007, Y. Saeys, I. Inza, and P. Larranaga proposed applying a feature selection technique to biological information, which was published in Bio-informatics and provides benefits and overviews of different feature selection technique for classification[5]. In the same year, D. Nadeau and S. Sekine uses Language factor, entity type factor, word level features, list look-up features, supervised learning, unsupervised learning to recognize information units like name of person, organization or location and numeric expressions like time, date, money and percent expressions[9].

In 2012, deep convolutional neural network is used by A. Krizhevsky, I. Sutskever, and G. E. Hinton to classify high-resolution images into the 1000 different classes. It contains five convolutional layers, some are max-pooling layers, and three fully-connected layers with a final 1000-way soft-max [7]. In 2014, Y. Kim used Word2vec, hyper parameter tuning and pre-trained word vector to use CNN on top of pre-trained word vectors for sentence-level classification tasks and allow for the use of both task-specific and static vectors[6]. In the same year, R. Patil applied Jelinek-mercer smoothing method and Bayesian model to predict heart diseases. It contains medical profiles such as age, sex, blood pressure and blood sugar, chest pain, ECG graph etc. it can predict the likelihood of patients getting a heart disease[8].

In 2015, Y. Lin, L. Huang, and Z. Wang used Intelligent medical guidance,TF-IDF algorithm, Cosine similarity to calculate possibility of disease for patients and gives the reasonable user guidance[10]. In 2016, Cuili Yao, Yue Qu, used Convolutional neural network and named entity recognition to predicts the diseases from patient’s symptoms and questions[1]. In the same year, Sunil kumar sahu and Ashish anand used Recurrent neural network and character level word embedding to recognize disease name and classify them into pre-defined categories and uses CNN to get character based embedded features[2].

Data Mining refers to using a variety of techniques to identify information or decision making knowledge in the database and extracting these in a way that they can put to use in areas such as decision making, predictions, for valuable forecasting and computation. This research work has developed a Decision Support in Heart Disease Prediction System (HDPS) using data mining modelling technique, namely, Naïve Bayes. Using medical profiles such as age, sex, blood pressure and blood sugar, chest pain, ECG graph, etc. It can predict the likelihood of patients getting a heart disease[13].

III. RELATED WORK

A. An intelligent medical guide system :

It is based on TF-IDF algorithm, the system framework as shown in figure 1, it consists three modules: 1) UI module, this module include user register, login, input symptoms and display medical guide results. 2) NLP (Nature language symptom) module, the module is the first segment the natural language symptom which is input by user. Word segmentation uses the IK Analyzer which is open-source and based on java. 3) MGC (Medical guide calculation) module, is the core module of this system[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)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

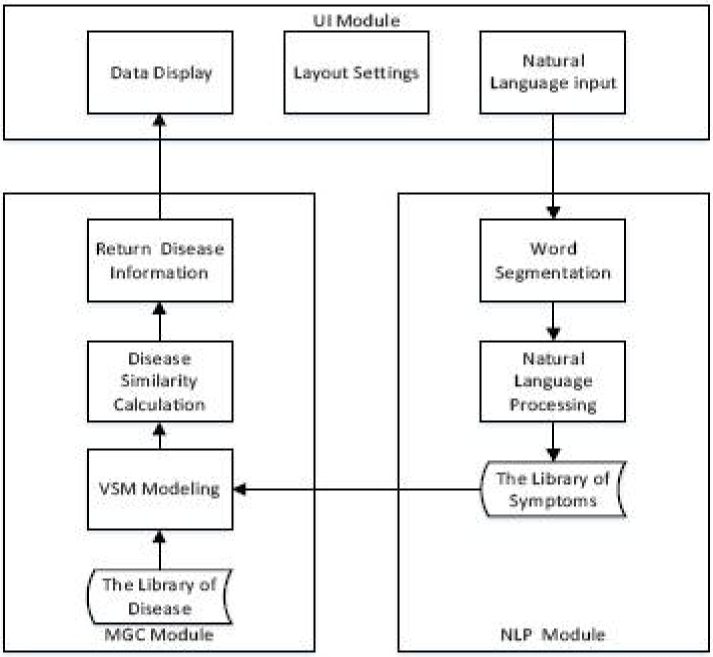

Figure1. System Framework[4]

TF-IDF algorithm is a method to calculating weight, is a kind of effective method always used in information retrieval and data mining. proposed to use the TF-IDF with semantic information to calculate similarity, proposed to use TF-IDF to identify user's hand writing, proposed to use TF-IDF to calculate similarity in the radar intelligence analysis, etc[4].

B. Modified Full Bayesian Classifier and Artificial Bee Colony:

The hybrid approach called Modified Full Bayesian Classifier and Artificial Bee Colony are used for medical diagnosis support system[11].

Phase 1: The data consists of medical history, symptoms, laboratory analysis … etc.

Phase 2: The patient may recall all symptoms for his diseases but the doctors selected only the main symptoms for obtained the batter diagnosis. The first step of filtering is to eliminate the redundancy of data by identifying the patient's records that have the same attributes and unified it's in one record. The last attribute in each table presents the number of patients that have same attributes. Second step of filtering is to delete all the attributes that presents secondary symptoms and medical history.

Phase 3: After filtering and extracting for data is stored all tables in repository of feature DB.

Phase 4: All dataset becomes input to M-FBC model. Checked the table in DB is ascertained to be empty and calculate (mutual information, threshold, weight and modified CPT (Condition Probability Tables) for each variable and modified Probability for structure).

Phase 5: Take each row from tested dataset and make test for Bns (Bayesian Network) by calculate probability base and calculate accuracy for the structure by equation

$$\text{Accuracy}(\%) = (\text{True} / \text{True} + \text{False}) * 100$$

Phase 6: When Complete all results creating fitness array, choose the highest accuracy from it's under in the fitness array can specify the number of columns in the matrix FSM (Food Source Memory) required and taken all the weights

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

in the specify column and make a new structure represents perfect solution (optimized) and calculate CPTs for each node in the structure and store it. Note that each table in dataset has its own structure can see the structure[11].

IV. PSEUDO CODE

C. Naïve Bayes and Jelinek-mercer smoothing:

The Naïve Bayes Classifier technique is mainly applicable when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods. Naïve Bayes model recognizes the characteristics of patients with heart disease. It shows the probability of each input attribute for the predictable state[13].

The naive Bayesian classifier, or simple Bayesian classifier, works as follows:

- Let D be a training set of tuples and their associated class labels as Ca and Cp. As usual, each record is represented by an n-dimensional attribute vector, X=(x1, x2…,xn-1, xn), depicting n measurements made on the tuple from n attributes, i.e. A1 to An.
- Suppose that there are m number of classes for prediction, C1, C2… Cm. Given a record, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naïve Bayesian classifier predicts that tuple x belongs to the class Ci if and only if

  $$P(Ci|X) > P(Cj|X) \qquad \text{for } 1 \le j \le m \text{ and } j \ne i$$

  Thus we maximize P(Ci|X). The class Ci for which P (Ci|X) is maximized is called the maximum posteriori hypothesis. By Bayes' theorem

  $$P(Ci \mid X) = P(x \mid Ci)\, P(Ci) / P(x)$$

- As P(X) is constant for all classes, only P (X|Ci)* P (Ci) need be maximized. If the class prior probabilities are not known, then it is often assumed that the classes are equally likely, that is, P (C1) =P (C2) =…P(Cm-1)=P (Cm) and therefore maximize P (X|Ci). Otherwise, maximize P (X|Ci) P (Ci). Note that the class prior probabilities may be estimated by P (Ci) =|Ci, D|/|D|**,** where |Ci, D| is the number of training tuples of class Ci in D.

- Given data sets with many attributes, it would be extremely computationally expensive to compute P(X|Ci). To reduce computation in evaluating P(X|Ci), the naïve assumption of class conditional independence is made. This presumes that the values of the attributes are conditionally independent of one another, given the class label of the tuple (i.e., that there are no dependence relationships among the attributes). Thus,

  $$P(X|Ci) = \Pi^{m}_{k=1} P(xk \mid Ci) = P(x1|Ci) * P(x2|Ci) * \ldots P(xm|Ci).$$

  We can easily estimate the probabilities P (x1|Ci), P (x2|C i)… P (xm|Ci) from the database training tuples. Recall that here xk refers to the value of attribute Ak for tuple X. For each attribute, we will see that whether the attribute is categorical or continuous-valued. For instance, to compute P (X|Ci), we consider the following:

  (a) If Ak is categorical, then P (Xk|Ci) is the number of tuples of class Ci in D having the value xk for Ak, divided by |Ci,D|**,** the number of tuples of class Ci in D.

- In order to predict the class label of X, P(X|Ci)P(Ci) is evaluated for each class Ci. The classifier predicts that the class label of tuple X is the class Ci if and only if

  $$P(X|Ci)P(Ci) > P(X|Cj)P(Cj) \qquad \text{for } 1 \le j \le m,\ j \ne i$$

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

In other words, the predicted class label is the class Ci for which P (X|Ci) P (Ci) is the maximum.

Steps to implement Naïve Bayes with Jelinek-mercer smoothing:
- Enter patient record.
- The two classes in which we have to classify the data are: 0:HD absent ,1:HD present
- Decide the probability of each attribute for both the classes using the database with result as training.
- Use Jelinek mercer smoothing for calculating smoothed probability of that attribute.
- In Jelinek-mercer smoothing the smoothing agent is used whose value ranges from 0 to 1.
- The value used in implementation is set optimally to give better performance for naïve bayes.
- Calculate the maximized probability from both the classes.
- Decide the class for patient record.

D. Convolutional Neural Network (CNN):

In the field of natural language processing, CNN reveals good performance in sentiment analysis, spam detection, and topic classification, and research on CNNs are continually emerging. The goal of CNN is to extract the symptom description from the user query and combine the symptoms to predict the most likely disease being suffered by patients under the combination of symptoms. The CNN network structure for disease classification and prediction is as follows[6].

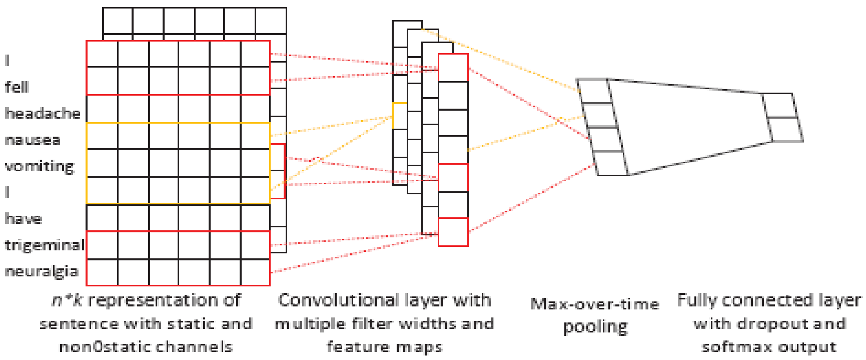

Figure2. Architecture of CNN for Text Classification[6]

Let $X_i \in R^k$ be a k-dimensional vector of the i-th word in the corresponding sentence. A sentence of length n can be expressed as X1: n = {X1 + X2 + …. + Xn} scenario, Xi : i+j is used to represent the concatenation of Xi, Xi+1, ….., Xi+j .The convolutional operation contains a filter $w \in R^{hk}$, and this filter is used to operate on the sliding window of length h to generate a new feature. For instance, a sliding window Xi : i+h-1 generates the feature ci according to the following equation:

$$Ci = f(w . xi : i+h-1 + b)$$

where b € R is an offset term and *f* is a non-linear function such as the hyperbolic tangent function. This filter is applied to all possible windows in the sentence, {X1:h X2:h+1 …. Xnh+1:n}, to generate the characteristic mapping c = [c1, c2 ….,cn-h+1]. Here, we have c € Rn-h+1. Next, each characteristic mapping is applied with a maximum pooling operation, where the specific operation is to take C^ = max{c} as the eigenvalue corresponding to the characteristic mapping. The goal of this operation is to extract the most important feature for each characteristic mapping, namely, the feature of maximum value. Each filter will generate a feature. This model applies the different filters to windows of varying size to generate multiple eigenvalues.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

These eigenvalues constitute the penultimate layer and pass a fully connected softmax layer. The output of this layer is the probability distribution on each classification label. In this paper, we select the diseases with the five highest probabilities and take the ratio between their probability and the probability of these five diseases as the predicted disease for output. the equation for calculating the precision is

$$precision = | correct | / | predict |$$

where, | correct | is the number of correct elements, namely the intersection of the output result and the Standard result, and | predict | is the number of elements in the output results.

E. CNN with RNN:

In many recent sentence representation learning works, neural network models are constructed upon either the input word sequences or the transformed syntactic parse tree. Among them, convolutional neural network (CNN) and recurrent neural network (RNN) are two popular ones. In general, CNN is able to learn local response from temporal or spatial data but lacks the ability of learning sequential correlations; on the other hand, RNN is specialized for sequential modeling but unable to extract features in a parallel way.

To get benefit from the advantages of both CNN and RNN, the CNN is constructed on top of the pre-trained word vectors from massive unlabeled text data to learn higher-level represent ions of n-grams. Then to learn sequential correlations from higher-level sequence representations, the feature maps of CNN are organized as sequential window features to serve as the input of LSTM. The combined approach utilizes CNN to extract a sequence of higher-level phrase representations, and are fed into a long short-term memory recurrent neural network (LSTM) to obtain the sentence representation. The combined approach is able to capture both local features of phrases as well as global and temporal sentence semantics.

Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have emerged as two widely used architectures and are often combined with sequence-based models. The capability of CNN to capture local correlations of spatial or temporal structures, CNNs have achieved top performance in computer vision, speech recognition and NLP. For sentence modeling, CNNs perform excellently in extracting n-gram features at different positions of a sentence through convolutional filters, and can learn short and long-range relations through pooling operations. CNNs have been successfully combined with both sequence-based model and tree-structured model in sentence modeling.

The other popular neural network architecture Recurrent Neural Network is a class of artificial neural networks which utilizes sequential information and maintains history through its intermediate layers. RNN does connections between units that form a directed cycle. This feature creates an state that is internal to the network which allows it to exhibit dynamic temporal behavior. Unlike feed-forward neural networks, it uses their internal memory to process arbitrary sequences of inputs. This makes them applicable to tasks such as unsegmented connected handwriting recognition or speech recognition[2]. RNN is able to handle sequences of any length and capture long-term dependencies. To avoid the problem of gradient exploding or vanishing in the standard RNN, Long Short-term Memory RNN (LSTM) and other variants were designed for better remembering and memory accesses. Along with the sequence-based or the tree-structured models, RNNs have achieved remarkable results in sentence or document modeling.

LSTM:

Traditional RNN models suffer from both vanishing and exploding gradient. Such models are likely to fail where we need longer contexts to do the job. These issues were the main motivation behind the LSTM model. LSTM layer is just another way to compute a hidden state which introduces a new structure called a memory cell (c_t) and three gates called as input (i_t), output (o_t) and forget (f_t) gates[2]. The LSTM architecture has a range of repeated modules for each time

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

step as in a standard RNN. At each time step, the output of the module is controlled by a set of gates in R^d as a function of the old hidden state h_{t-1} and the input at the current time step x_t. These gates collectively decide how to update the current memory cell c_t and the current hidden state h_t . d is denoted by the memory dimension in the LSTM and all vectors share the same dimension[13].

Computation of memory cell (c_t) is done through previous memory cell and candidate hidden state (q_t) which we compute through current input and the previous hidden state. The final output of hidden state would be calculated based on memory cell and forget gate[2]. The LSTM transition functions are defined as follows:

$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$
$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
$q_t = \tanh(W_q \cdot [h_{t-1}, x_t] + b_q)$
$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$
$c_t = f_t \square c_{t-1} + i_t \square q_t$
$h_t = o_t \square \tanh(c_t)$

Here, σ is the logistic sigmoid function that has an output in [0, 1], tanh denotes the hyperbolic tangent function that has an output in [−1, 1], and □ denotes the element wise multiplication. To understand the mechanism behind the architecture, f_t can be viewed as the function to control to what extent the information from the old memory cell is going to be thrown away, i_t to control how much new information is going to be stored in the current memory cell, and o_t to control what to output based on the memory cell c_t . LSTM is explicitly designed for time-series data for learning long-term dependencies, and therefore we choose LSTM upon the convolution layer to learn such dependencies in the sequence of higher-level features[13].

V. CONCLUSION AND FUTURE WORK

Neural network is one of the better approach used for on-line medical guidance. CNN is self learner and performed feature extraction giving high accuracy. It has been found from the survey of the previous work that instead of using only CNN for on-line medical guidance if CNN with RNN is used, the computation can be decreased and efficiency can be improved.

REFERENCES

1. Cuili Yao, Yue Qu, Bo Jin, Li Guo, Chao Li, Wenjuan Cui, and Lin Feng,“A Convolutional Neural Network Model for Online Medical Guidance”, Institute of Electrical and Electronics Engineers (IEEE), pp. 4094 – 4103, July 2016.
2. Sunil Kumar Sahu and Ashish Anand, “Recurrent neural network models for disease name recognition using domain invariant features”, Association for Computational Linguistics, pp. 2216–2225, june 2016.
3. R. Ledley and L. Lusted, “Reasoning foundations of medical diagnosis”, MD computing: computers in medical practice, Vol. 130, pp. 9-21, 1991.
4. Y. S. Lin, L. Huang, and Z. Wang, “An intelligent medical guidance system based on multi-words tf-idf algorithm,” in Energy Science and Applied Technology: Proceedings of the 2nd International Conference on Energy Science and Applied Technology (ESAT 2015). CRC Press, pp. 385–389, 2015.
5. Y. Saeys, I. Inza, and P. Larra˜naga, “A review of feature selection techniques in bioinformatics,” bioinformatics, vol. 23, no. 19, pp. 2507–2517, 2007.
6. Y. Kim, “Convolutional neural networks for sentence classification,” Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 1746–1751, 2014.
7. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, pp. 1097–1105, 2012.
8. R. Patil, “Heart disease prediction system using naive bayes and jelinek-mercer smoothing,” Int J Adv Res Comput Commun Eng, vol. 3, no. 5, pp. 701-708, 2014.
9. D. Nadeau and S. Sekine, “A survey of named entity recognition and classification”, Lingvisticae Investigationes, pp. 303-326, 2007.
10. Y. Lin, L. Huang, and Z. Wang, “An intelligent medical guidance system based on multi-words tf-idf algorithm,” in Energy Science and Applied Technology: Proceedings of the 2nd International Conference on Energy Science and Applied Technology (ESAT 2015), pp. 385–389, 2015.

11. A. T. Sadiq and N. T. Mahmood, "A hybrid estimation system for medical diagnosis using modified full bayesian classifier and artificial bee colony", Iraq Academic Scientific Journals, ISSN: 00672904, pp. 1095-1107, 2014.
12. B. Settles, "Biomedical named entity recognition using conditional random fields and rich feature sets," in Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications, Association for Computational Linguistics, pp. 104–107, 2014.
13. Chunting Zhou, Chonglin Sun, Zhiyuan Lui, Francis C.M. Lau, "A C-LSTM Neural Network for Text Classification" no. 10, 2015.

BIOGRAPHY

Madhuri N. Suryatal is a student of M.E. in the Computer Engineeering Department, Pune Institute of Computer Technology. Her research interests are Deep Learning, Machine learning, Data Mining.

P. S. Vidap is a professor at Department of Computer Engineering, Pune Institute of Computer Technology. Her research interests are Deep Learning, Machine learning, Data Mining.