

# A SURVEY ON IMPORTANCE AND APPLICATIONS OF DATA MINING USING ARTIFICIAL NEURAL NETWORK

Ms. R.Jaya, Mr.Giridhar V

Sr. Assistant Professor, Department of Computer Science and Engineering

U G Student, Department of Computer Science and Engineering,

New Horizon College of Engineering, Bangalore, Karnataka, India

[jayamanojkumar@gmail.com](mailto:jayamanojkumar@gmail.com), [giridharv10@gmail.com](mailto:giridharv10@gmail.com)

**ABSTRACT**-As there have been increased modernisation curve in development field of each and every equipment, process or technique involved in modern medicine world. In Data mining pattern recognition and pattern extraction are interpreted, explored, realised and analysed to finally evaluate, arriving at expected conclusion having uplifting accurate and respective diagnosis to respective cause. DM techniques with utmost focus on NN and ANN as well as DM tasks of conceptual inference in mining and matching are imprinted in this paper

Keywords - DM techniques, DM tasks, ANN, NN, Association rules, Naive bayes, Clustering

## I. INTRODUCTION

An artificial intelligence mechanism exercising computational practices exploring patterns to predict essence of recognized patterns in particular field for a particular cause with a brief analysis from statistically stored database catalogues leads to current domain studies in data mining, adaptive flourishing domain of computer science in medical field. The normal lifestyle, adulterate food, habits and habitats etc., has given verdict to reduced lifespan. Data mining has built up statistical analysis of exaggeration and success, building the domain from the scratch with new conceptualization of disease cure.

**Importance of data mining:** Data mining is core knowledge from where essential data is captured or mined for some basis .its gaining demand in Business transactions, Scientific data analysis, Medical research and personal data scrutiny, Surveillance video and pictures queries and also in field of Satellite sensing, Games, Virtual Worlds. We extract the required data in a substantial manner from databases indulging in cleaning database, integrating elements of designated data to form data warehouses further selecting and transforming to task related data mining involving pattern extraction to gain knowledge[16][9].

Data mining is an interdisciplinary subfield in computer science domain with computational process of discovering patterns in big data involving collective methods of artificial intelligence, machine learning, statistics, analytics etc., Data mining involved with fetching data from enormous warehouses imply task querying called data mining task primitives, these primitives form interactive communication during discovery in order to direct the mining process, or examine the mining deeds from different angles or depths.

### Related research work

World towards advancement has been improving technology in every domain but still lifespan of humans is not meeting up to the mark as it's expected to be, according to bio science. The present lifestyle, threatening to be the cause of reduced lifespan. There has been series of heart, kidney, and

lung etc., disorders being encountered [1] echo's the diagnosis of heart disease using predictive DM where the survey report of heart patients is collected and diagnosis is involved with Naïve Bayes and Decision Tree algorithms. The prediction of disease is by DM weka tool where dataset containing attributes and values forms input and heart patients prediction forms output. The brief discussion about dataset, tools and principles involved in predictive DM is emphasized.

The naive bayes algorithm has imprinted a new channel of exploration in understanding data[6]. Conveys the principle, implementation, objectives, rules and step by step algorithm analysis of naïve algorithm along with Jelinek-mercer smoothing technique[2]. Intelligent heart disease prediction system mechanism envelopes analysing, predicting heart diseases based on person's lifestyle and symptoms he suffers from. The system which is expandable incorporating numerous records or attributes with significant rules under DM, also extracts masked essential historic data from medical database. jelinek smoothing algorithm has more weighed thrust in prediction answering complex queries with a flexible move, providing ease of access detailed information and accuracy[7].

Technology dependency on bio system or bio system on technology is incredible. Introducing chemical kinetics of life, modelling kinetics of drug release, classification and species determination is made through advanced computational technologies as ANN, briefly discussing the neural network model with mathematical analysis showing network learning and database training structures, made clinical context as input for ANN to respond giving out diagnosis related output, explaining overviewed picture of cancer, cardiovascular and diabetes diseases. Finally describing fundamental steps in ANNs-based medical diagnosis involving features such as -Building the database Data cleaning and preprocessing Data homoscedasticity - Training and verification of database using ANN Network type and architecture Training algorithm Verification Robustness of ANN-based approaches -Testing in medical practice etc., are depicted[8].

The improvement in field of science is has made us to march towards discovering new diseases classifying them to get a cure. Artificial intelligence based expert system gobble up entire catalogue database of crucial cases critically, to provide taxonomy depicting essential description. An expert system (ES) known as knowledge based system, is a computer program that uses knowledge and inference procedures to solve problems that are ordinarily solved through human expertise. It has mainly three components a) knowledge base, b) inference engine c) user interface. The diagnosis of coronary heart diseases (CHD) is dealt along with proposed framework of algorithmic descriptions and rule set of CHD such as Angina pectoris, Myocardial infarction(heart attack) etc., are focused[9].

Renewed interest in the 2000s in Neural Networks was sparked because of the advent of deep learning. In CMOS, Computational devices have been created, to both neuromorphic computing and biophysical simulation. In the computational devices had big improvement from 2006 [11]. In this paper [14] the author tells us that in data mining techniques neural network is widely used. Multilayer perception is very slow for the high dimensional datasets. In real time data mining the application it is an important drawback for using them, because its requirement is that, it needs answer within a short duration of time [13]. CMAC Neural Network adaptation they have proposed for data mining in this research paper. Which provide guaranteed convergence and fast training time? This paper also defines us that how CMAC adaptation is constructed for data mining, about classification model that is used in data set of real-life. It works on real dataset. The models which were used are CMAC, MLP and C4.5 and the algorithm is CMAC algorithm [13].

In this paper [15] the author tells us about a new method in which DSS is supported based on data mining and neural network. Three subjects have been discussed in this paper. To support DSS how to make use of neural network is the first. And the second one is that, to support DSS how to make use of data mining and the third one is, To support DSS how to make use of neural network and Data Mining. DSS was developed by American scientist Scott Three departments of DSS are discussed they are Knowledge system (KS), and Problem process system (PPS) and language system (LS). The neural network to DSS function is ensuing by scrutinizing the characteristic of problems and neural network of DSS. The large scale parallel distribution processing function, learning function, selflearning of natural language processing system , overcoming the difficulties of infinite recursion and assembled blast, promoting of deciding the ability of DSS and real time processing , Adoptive parallel association reasoning [15].

## II. CATEGORIES OF DATA MINING

Data mining tasks are of two categories: Descriptive data mining and predictive data mining, where descriptive DM describes information of interest without predetermined analysis. Predictive DM analyse to fetch result in predictive manner by allowing unknown field values to be submitted,

followed with interpreting patterns encountering output information with pattern mapping experience.

DM categories involve classification, prediction, clustering and association rules, the former two categories descend from prediction DM and latter categories descend from description DM[6][9].

**Classification:**The primary step under focus in DM is to classify by projecting different items to their respective groups through pattern recognition skill and experienced examination. Classification is supervised learning as data classes or concepts are predefined describing even the characteristics of a set of data classes.

**Prediction** is of construction, estimation, perception, detection along with effective utilisation of predicted one's. it tracks the class of unlabeled object or needed value from varied range frequency, followed with forecasting to generate the result of accuracy by estimating future value of continuous variables based on live patterns within the study.

**Clustering** is unsupervised learning process which envelope classification techniques, with no group predefinition. The key concept of clustering is to group abstract object based on quantitative similarities. Clustering is not convergent to specific algorithm but its general and divergent[6][16].

**Association rules** are being used in research which aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories having communicative effective interface associating matching protocols[1][16].

**Neural Networks:** The inspiring traits for neural networks being animal behaviour responding to stimuli. It's one of the best algorithms in prediction and diagnosis of illness. Cognitive science based NN, An artificial intelligence based data mining technique being made reputation in healthcare technology.

The input information is perceived by dendrites of neuron generating a non linear response when adequate threshold information is reached and collected at synapse to communicate and pass other neurons via axon. All attributes values must be encoded in standardized manner .This technique works well as long as minimum and maximum values are known and all new potential data at reference are in boundary. NN are robust to minor disturbances if violations exceed control certain ad hoc solutions or extensions of boundary will be undertaken.

**NN algorithm:** it works through training data set neurons record by record adjusting weights constantly reducing prediction errors. It involves training through number of passes rendering conclusion at required diagnosis. Training period if reduced affects collapsing the models efficacy. As NN are prone to over fittings as they flexible enough causing threshold to reach less than minimum criteria. The impact of abrupt termination is catastrophic; we avoid such termination by cross validation check procedure involving five steps.

1. To retain part of the original data set as a holdout validation set.
2. Proceeding to train neural network as above on remaining training data.

3. Apply the weights learned from the training data on the validation data
4. Monitor two sets of weights, one current set of weights produced by training data, and one best set of weights as measured by the lowest Sum Squared Error so far on the validation data.
- 5 When the current set of weights has significantly greater SSE than the best set of weights

**ANN's:** A mathematical representation of human neural architecture, reflecting its learn content and skill full knowledge with generalisation capabilities. It belongs to artificial intelligence domain, imprinting research related models, echoing high level non-linear mechanism patching complex relationships among the variables describing threshold weight to data mine critical information to be popped out at required times. Moreover, ANN is an adaptive system that changes structure of threshold weighted communicative neural mechanism based on inputs that enter during learning phase of training.

A neural network is giant clusters of numerous neurons which are framed to fit in an organized layer view. The weighted connection confirms the neural sensitive data transfer approach to frame ANN methodology. The structure of neural network is formed in between layers, having a layer of inputs giving raise to next set of hidden layers to march out at output layer.

ANNs are at more fame by their accuracy, has made footstep in replacing traditional diagnosis and classification approaches. they are nonlinear, nonparametric, adaptive and can theoretically approximate any fundamental relationship with arbitrary accuracy [3].

Choosing ANNs in diagnosis has its importance for these generalised reasons as an iterative training approach involved in analysing implemented solutions ,simple architecture for physical implementations, the feature of generalising input vectors, complex class distribution mapping ,ANNS ability to be implied as arbitrary function approximation mechanism that grasps from observed data. ANN deals with function approximation problem which involves selection of a specific function among a well-defined class to closely match a target function in a task-specific way.

**Working of ANN:**

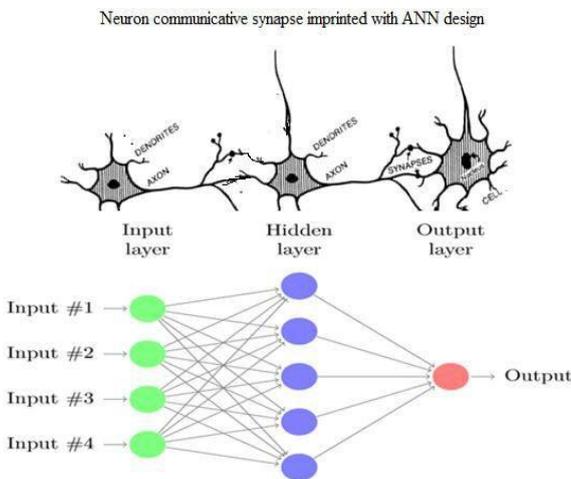


Fig 2: Neuron communicative synapse imprinted with ANN design

Reticulate structure of intricate neurons clustering dynamically, interactively in a self organised way to process the information of interest.‘ Integrated circuit and advanced technology with two dimensional devices having limited layered interconnections \_of fact are brought to physical reality considering restriction to types, scope of artificial neural networks while implementing with silicon. Neural networks of current medicinal technology are simple clustered web of primitive artificial neurons .layered clustering mechanism forming intricate network to resolve biomedical problems in optimised procedure is thoroughly through engineered human toil of glide.

Generalised similar neural network layered structure or topology for many medical cases has been employed. The topology has some neuron interface to experience real circumstance in fetching inputs .few other neurons provide the real world with the network's outputs. The output being the representation of input, handling the reflex arc, the output view has scanned mechanism of patterning the image of importance generating response to stimuli. The sandwiched neural layers \_between outer and inner layers are hidden to sight.

Earlier research was made connecting neurons at random manner, which remain false positive towards success. Even the primitive invertebrates considering from mollusc phylum have structured neural ganglia or brain, then enhanced research incorporating layered network approach, summation and transfer function involved promoting to compute in computer functional neural networks procedure. The lines of communication during transactions between neurons form the thesis of neural network interactive approach. The package wrapping selfdom communication between layers form the asset of neural networks, implying variable input strength of neural networks giving out result of necessity. There are two communicative behaviour of connections one being summing information to neighbouring neuron layer called -excite behaviour and another one being subtractive information of no advancement to next neuron called -inhibit behaviour. The probability of pattern match should yield approximate result ceasing confusion leading to important case called lateral inhibition. Neurons of same layers when dealing with probabilities avoid confusions called lateral inhibition. The network chooses the highest probability inhibiting all others using lateral inhibition. This conceptualisation of lateral inhibition is also called competition.

Connections are of feed forward pathway and feedback pathway, the former approach is forward communication considering from input to output layer and in the latter approach the output of one layer routes back to a previous layer. The way that the neurons are connected to each other has a significant impact on the operation of the network. The implementation of professional software package to schedule controls such as add, delete etc., coordinates to excite or inhibit behaviour as employed in neural study.

The first step towards Working of ANN is signed with training ANN, the structure for an application once prepared to be functioned then mechanism is triggered towards training. The loads of training expressed with varied weighed analysed input note or data are chosen randomly at initial circumstance, followed with learning and training.

There are two approaches to training - supervised and unsupervised.

In supervised training both inputs and outputs are decided which are provided for processing. The network then processes the inputs and compares its resulting outputs against the optimised expected or desired output. If errors or exceptions discovered the corrective procedure is triggered to back propagate all through the system re-correcting encountered flaws in communicative mechanism. Back propagation has been tedious, as it needs to locate capturing point of flaw and correct it by adjusting the weights of input which control the network. This process is being iterative as it's involved with adjusting varied range of weights, allowing the neural data of answered result to flow. The set of data which enables the training is called the "training set." During the training of a network the same set of data is processed many times as the connection weights are ever refined. The current commercial network development packages provide tools to monitor how well an artificial neural network is converging on the ability to predict the right answer focusing optimal view. These tools allow the training process to be continuously active for some days until the system reaches some statistically desired point of accuracy predicted.

To monitor the network to determine if the system is simply memorizing its data in some non significant way, supervised training needs to hold back a set of data to be used to test the system after it has undergone its training. New adaptive training techniques are fed based on necessity. When finally the system has been correctly trained, and no further learning is needed, the weights can, if desired, be "frozen."

Unsupervised training: the model is not provided exactly the ordered sequence of inputs and outputs; the procedure followed during training is not step by step. The network is provided with inputs but not with desired outputs. The system itself must then decide what features it will use to group the input data. This is often referred to as self organization or adaption. This adaption to the environment is the promise which would enable science fiction types of robots to continually learn on their own as they encounter new situations and new environments. This unexpected aspect to life and the human desire to be prepared, there continues to be research into, and hope for, this field. Based on statistical characteristics or properties only we cluster input data in classes. Cluster significance and labelling is employed. The labelling can be carried out even if the labels are only available for a small number of objects, representatives of the desired classes[2][3].

**Application of neural networks:** Function approximation, regression analysis, time series prediction, Call control, sequence recognition, novelty detection, sequential decision making, Data processing, Language Processing, Character Recognition, Image (data) Compression, Pattern Recognition, Signal Processing, Servo Control ,financial sector database etc.,[4]

**Application of artificial neural networks:** Load Forecasting, Fault Diagnosis\Fault Location , Economic Dispatch, Security Assessment, Alarm Processing, Eddy current analysis, Harmonic source monitoring, Load frequency control (Automatic Generation Control),

Hydroelectric Generation Scheduling, Power System Stabilizer Design ,medical diagnosis, financial application[5].

### III. CONCLUSION

At present data mining is a new and important area of research and ANN itself is a very suitable for solving the problems of data mining because its characteristics of good robustness, self-organizing adaptive, parallel processing, distributed storage and high degree of fault tolerance. The commercial, educational and scientific applications are increasingly dependent on these methodologies. This paper has provided the summary of one of data mining technique ANN used for medical data mining in the field of healthcare and also the diseases they classified. It also expresses the importance and challenges of data mining with respect to neural and artificial neural networks.

### REFERENCES

- [1] Sotiris Kotsiantis, Dimitris Kanellopoulos -*Association Rules Mining: A Recent Overview* Educational Software Development Laboratory Department of Mathematics, University of Patras, Greece
- [2] [http://www.astro.caltech.edu/~george/aybi199/Donalek\\_Classif.pdf](http://www.astro.caltech.edu/~george/aybi199/Donalek_Classif.pdf)
- [3] Application of data mining using artificial neural network :survey", *Intl journal of database theory and application*, vol 8 2015
- [4] <http://www.psych.utoronto.ca/users/reingold/courses/ai/cache/neural3.html>
- [5] Sonali Maing,Priyanka Wankar, "Research paper on basic of artificial neural network ",*Intl Journal of recent and innovation trends in computing and communication vol2* issue 1 jan2014
- [6] R. Jaya, S. Mohan Kumar -A Study On Data Mining Techniques, Methods, Tools And Applications In Various Industries| *Int J Current Research Rev / Vol 8 • Issue 4 • February 2016*
- [7] Filippo Amato, Alberto López, et al "Artificial neural networks in medical diagnosis ",*Journal of applied biomedicine 2013* issn 1214-0287
- [8] R.A. Soltan M. Z. Rasha, B.El-Desouky, "Diagnosis of Some Diseases in Medicine via computerized Experts System" vol 5, october 2013 *International Journal of computer science and information technology*
- [9] R. Jaya, S. Mohan Kumar| Ayurveda Medicine Roles in Healthcare Medicine, and Ayurveda towards Ayurinformatics| *IJCSMC, Vol. 4, Issue. 12, December 2015, pg.35 – 43*
- [10] Nikita Jain , Vishal Srivastava "DATA MINING TECHNIQUES: A SURVEY PAPER " *IJRET: International Journal of Research in Engineering and Technology* eISSN: 2319-1163 | ISSN: 2321-7308 Volume: 02 Issue: 11 | Nov-2013
- [11] Rasika P. Ghom , N. R. Chopde" Survey Paper on Data Mining Using Neural Network" *International Journal of Science and Research (IJSR)* ISSN (Online): 2319-7064

- [12] Mariammal,Dayanthi,Dr.Patra, "Major disease diagnosis and treatment suggestion system using data mining techniques", *IJARCSST 2014,VOL2 JAN-MACH 2014*
- [13] F. Palacios, X. Li and L. E. Rocha, -Data mining based on CMAC neural networks, Electrical and Electronics Engineering, 2006 3rd International Conference, IEEE, (2006), pp. 1-4.
- [14] T. Nagabhushan, H. Ko, J. Park, S. Padma and Y. Nijagunarya, -Classification of Symbolic Objects Using Adaptive Auto-Configuring RBF Neural Networks, Information Technology Convergence, 2007. ISITC 2007. International Symposium IEEE, (2007), pp. 22-26
- [15] X. Qian and X. Wang, -A New Study of DSS Based on Neural Network and Data Mining, E-Business and Information System Security, 2009. EBISS'09. International Conference IEEE, (2009), pp. 1-4
- [16] R. Jaya, S. Mohan Kumar -A Survey On Medical Data Mining Healthcare Related Research And Challenges” *International Journal of Current Research Vol. 8, Issue, 01, pp.25170-25173, January, 2016*



R.JAYA is working as a Sr.Assistant Professor in the department of CSE at New Horizon college of Engg.She has 13 years of teaching experience. She has done bachelor and master degree in engineering under Anna University. She is doing Ph.D under VTU. Her research interest includes data mining, expert systems,SOA and security.