

An Experimental Study for Analysis Approaches in Predicting the Risk for a Disease using Supervised Machine Learning Classifiers

Giddaluru Lalitha, Burra Krishna Sai, V Pavan Kumar

lgiddalu@gitam.edu, Krishnasaiburra1999@gmail.com, pavankumar.v.akt@gmail.com

1,2,3Department of Computer Science and Engineering, School of Technology, GITAM (DEEMED TO BE UNIVERSITY), Hyderabad, Telangana, India.

Abstract

Data Mining is a process which discovers interesting patterns from the huge amount of data. A computer learns of an activity without involving humans using Artificial intelligence with the help of an application called Machine learning. A science allowing computers to learn and act very much alike humans by accessing the data and other necessary information without any program is called Machine learning. Applications uses Machine Learning algorithms to optimize speech recognition etc. A Fastest growing areas in a computer science with long term approach applications is Machine learning which also refers to the automatic detection of noteworthy patterns in data with machine learning tools, which gives the ability for a program to learn and adapt. It has become one of the main pillar in IT and, which is hidden part of our life. In technological process the intelligent data analysis will spread with an increasing availability of data. This paper concerns the automated identification of intrauterine growth restriction (IUGR) types by the use of machine learning methods. In Medicine, there are many diseases and diagnoses where identification of their subtypes affects medical treatment. This paper discusses data mining techniques, Machine learning techniques, it's applications and Case Study on Disease Prediction

Key words : Data Mining techniques, Machine Learning Techniques, Supervised learning, Applications.

I. INTRODUCTION

This Data Mining is an emergent technology that is adopted by a large number of organizations. Because companies maintain large amounts of data, which is in an unprocessed format, data mining techniques are used to convert the bulk of data into golden nuggets of knowledge. The main goal of data mining is to find interesting patterns that are used for making decisions and improving the business of the organizations [1]. Database technology progressed from Primitive file processing into a Database management system, Advanced databases, Data mining and data warehouse. Data mining is also called as data dredging, Knowledge discovery from data or Knowledge extraction.

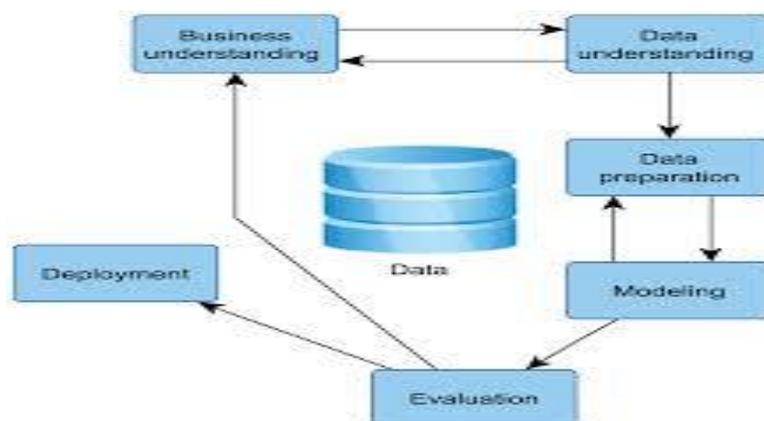


Fig.1. Ontology Design of Data Mining

Ontology data mining explains the framework for data mining which includes data mining entities like data sets, data tasks, definitions, algorithms, components. It also allows for complex concepts such as constrained based mining, queries, etc. The ontology design of data mining consists of 6 phases.

1. **Business Understanding:** Understanding the Business
2. **Data understanding:** Understanding the data according to the Data mining task.
3. **Data Preparation:** Preparing the data means performing pre-processing techniques on the data like cleaning, integration, transformation, reduction.
4. **Modeling:** Modelling the data according to requirements and data mining tasks.
5. **Evaluation:** Evaluation means determining the important variable, statistics and data, based on the problem.
6. **Deployment:** Interesting patterns are presented to the user[2].

II.UNITS

2.1 Relationship between Data Mining and Machine Learning

Machine learning is a process or collection of algorithms that improve their performance. Machine learning is used where humans do not exist. For example, navigating on mars where humans can't show their expertise (speech recognition), so the models must be customized. Machine learning methods have been used to train computer-controlled vehicles to steer correctly when driving on different types of roads. Mainly machine learning performs tasks such as Identifying anomalies, identifying patterns, Generating patterns, Prediction.

Identifying anomalies: Recognizing errors such as unusual bank transactions like a credit card.

Identifying patterns: Identifying patterns such as facial expressions, spoken words and medical images.

Generating patterns: Generating patterns like images or pictures.

Prediction: Estimations like future stocks and currency values.

2.2 Data mining Tasks:

Data mining tasks are classified into 2 ways namely Descriptive Mining task, Predictive Mining task.

Descriptive Mining Task: A task which enables descriptive mining to determine patterns and relationships in sample data. Typically, a descriptive model is found through aimless data mining. i.e. a bottom-up approach where the data “speaks for itself”. Aimless data mining finds patterns in the dataset but leaves the interpretation of the patterns to the data miner.

Predictive Mining Task: predicting the data values by making use of known results from a different set of sample data is called predictive mining. From the given know variable values including previous target variable values the predictive model is created. The training data consists of measurement pairs where each consisting of an input vector $x(i)$ along with a corresponding target value $y(i)$. The estimation of the function $y=f(x; q)$ which can predict a value y , given an input vector of measured values x and a set of estimated parameters q for the model f is called predictive model. The process of finding the best q values are the core for the data mining technique [3].

Data Mining Techniques: To specify the kind of patterns to be found in data mining where its functionalities are being used.

A. Discrimination and Characterization:

A summary of the general features of the target data classes is called characterization. For example, characterizing general characteristics of students who got a distinction. The result could be a general profile of students such as hard workers, knowledgeable, etc. Discrimination means comparing the target class with one or more contrasting classes. An example is comparing two groups of students one of which has students who got distinction every semester and the other with one who failed every semester. The resulting description says that 89% of the students are regular , hard worker, clever whereas 20% of the students are irregular,dull,etc[4].

Table1: Data Mining vs Machine Learning

Factors	Data Mining	Machine Learning
Scope	Through data visualization and pattern techniques, find out how various dataset attributes are related to each other. The main purpose is to identify the relationship between 2 or more dataset attributes and predict the outcomes or actions.	It is used to make the forecasts of the outcomes such as approximate time duration or price estimation. With an experience over a period of time it automatically learns the model.
Working	In order to collect the useful information digging deep into the data technique is used.	By using a method of improving complex algorithms by iteratively feeding the machines with a trained dataset in order to make it nearly perfect is called machine learning.
Uses	A research field such as fraud detection, web mining and text mining is more often used in Data Mining	In making recommendations of products, prices, delivery time estimation, etc. are the major uses of machine learning
Concept	To extract information using	Machines learn from the

	techniques and find out the trends and patterns is the concept behind mining.	existing data and improves by it. To build models on the logic behind data which can predict the future outcome are uses of data mining methods
Method	At any particular point of time data mining performs the analysis in batch format rather than continuously.	It uses the data mining technique to improve its algorithms and change its behavior to future inputs. In order to analyses the occurrence of the failure machine learning algorithms will continuously run and improve the performance of the system automatically. Without the need to human involvement or any new program the machine will updated based on the new or change in the trends.
Nature	human intervention is must to retrieve any information by applying the different techniques.	Machine learning is done automatically without any intervention which is quite different from data mining.
Learning Capability	Before applying any techniques, a manual intervention is must, so it is also called as manual technique.	Machine learning is more accurate than data mining as it uses same techniques used by the data mining which indeed automatically learns and adapt to the changes. So, machine learning is more accurate than data mining.
Implementation	Implementation is involved in construction of new models by using data mining techniques. A CRISP-DM kind of model is built based on the same data mining technique. In order to discovery of knowledge for pattern evaluation, data mining engine and databases which are used by data mining process.	A Machine Learning algorithm is used in artificial intelligence, neural network, neuro-fuzzy systems and decision tree, etc. In order to predict the future machine learning uses neural networks and automated algorithms.
Accuracy	The data mining accuracy is completely dependent on how the data is being collected. It produces accurate results that are being used by machine in order to produce better results. Data mining may miss	Data mining techniques are less accurate compared to machine learning algorithms.

	important relationships as it requires human intervention.	
Applications	On lesser volume of data only data mining can produce the results compared to machine learning.	Only in standard format the data needs to be inserted for the working of machine learning algorithms which are limited. All data needs to be converted from native format to standard format which are from multiple sources being scrutiny by the machine learning. For accuracy in results a large amount of data is required.

Classification: Classification is one of the important techniques for finding the categorical class labels. Classification is finding a model like an algorithm that finds the unknown class labels. To identify the applicants based on their risk profile, an LIC officer needs to analyze her data based on class label's like risky and safe. Classification is two-step process. Namely, the Learning step and the classification step. In the Learning step training data is diagnosed by a classification algorithm and Classification step tests data which is used to estimate the efficiency of the classification rules. The classification technique can be described in three ways, such as IF-THEN rules, decision trees, neural networks.[6]

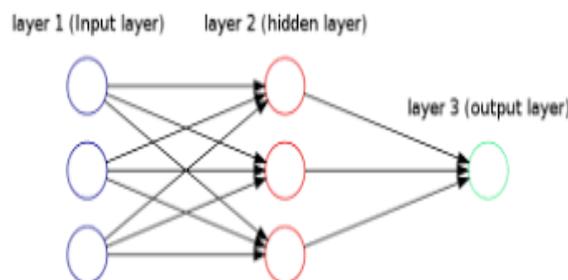


Fig: 2: Neural Network

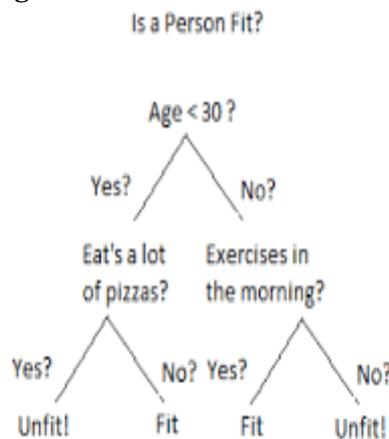


Fig.3. Decision Tree

These are the IF-THEN rules generated from the decision tree.

IF age<30=yes ^ eats lot of pizzas= yes => class label=Unfit

IF age<30=yes ^ eats lot of pizzas= No => class label=fit

IF age<30=No ^ exercise = yes => class label=fit

IF age<30=No ^ exercise = No => class label=Unfit

Classification Techniques:

Bayesian classification

Rule-based classification

Decision Tree Induction [6]

C.Prediction: Prediction is one of the important techniques which estimates the future data based on present data. Suppose consider the stock market. Customers will invest shares based on the present price and estimate the future price. Regression techniques can be used for prediction. Regression is one of the techniques used for predicting numeric class labels. Regression techniques are classified into Linear Regression, Non-linear Regression, Multivariate Linear Regression, Multivariate Non-linear Regression [7].

D. Clustering: Clustering is a process of grouping similar objects together. Objects which have similar properties are placed in one region otherwise are placed outside the region. We can identify sparse and dense regions in objects by applying different clustering techniques. And also we can discover overall common patterns and correlations among data attributes. Classification Algorithm used for finding the distance and similarities between clusters. For example, consider bank data, form a group which is taking education loan based on similar patterns.. There are 4 types of clustering methods namely Partitioning Techniques, Hierarchical Agglomerative (divisive) Techniques, Density-based Techniques, Grid-based Techniques.

E. Outlier analysis: Outlier means errors or exceptions. Analysis of noises is called outlier mining. An example is fraud detection. Outliers can be detected by using statistical methods and distance measures and also deviation-based Techniques.

F. Evolution Analysis: Evolution analysis represents models' patterns. This model is used for objects whose behavior changes according to their time. An example is the stock market. An evolution analysis study of stock exchange data and identify regularities. Such regularities may help estimate the future swings in stock market prices [10].

III. Data Mining used in various sectors

Data mining is used in various sectors. Most of the organizations using Datamining techniques include retail stores, hospitals, banks, and insurance companies, Intrusion detection etc. Many of these organizations are using things such as statistical techniques, pattern recognition, and other important tools. Data mining can be used to find patterns and important knowledge. This technology is very important in many business areas. because it allows them to learn more about their customers buying habits and make smart marketing decisions. Here is an overview of business problems and solutions found using data mining technology [10]. Real-time applications such as Risk assessment, Visual Recognition, Image classification Visual Recognition.

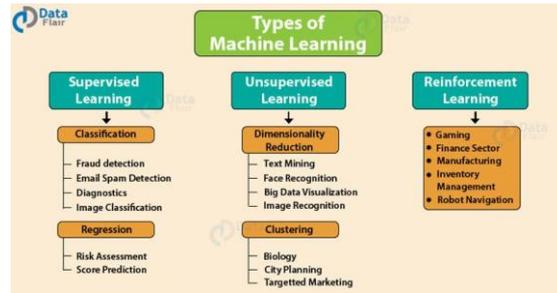


Fig.4. Machine Learning Algorithms Types

IV. Machine Learning Techniques

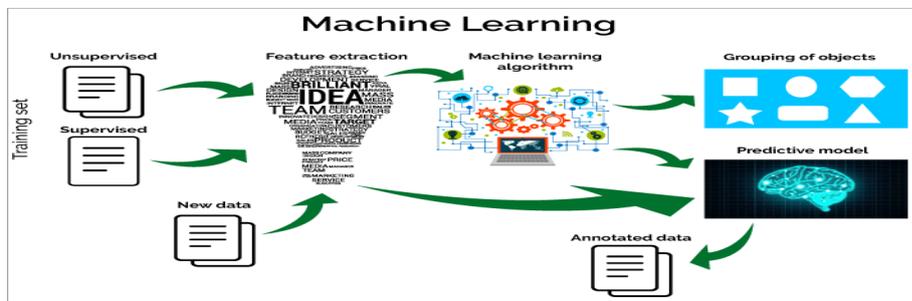


Fig.5. The Process of Machine Learning

Machine learning is one of the important and fastest-growing areas of computer science, which is used in many applications. It refers to the automated detection of meaningful patterns in data. Machine learning tools are concerned about providing programs with the ability to adapt and learn. Machine Learning (ML) is the science of getting computer programs to act without being specific programmed. Machine Learning techniques implemented in many areas such as self-driving cars, web search etc. Supervised classification is one of the tasks most frequently carried out by intelligent systems. Machine Learning is a technique that design complex algorithms for processing large bulk of data and delivers results to its users. It solves complex programs that can learn based on experience and according to that make predictions. Through regular input of training data these algorithms are improved by itself. The goal of machine learning is to analyze and understand data and construct models from training data that can be used by humans.

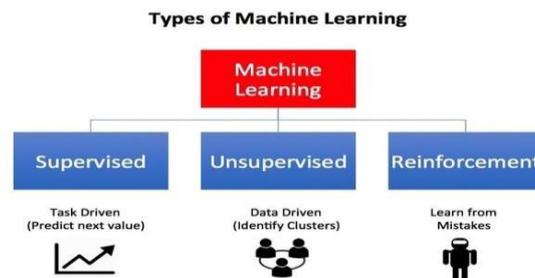


Fig.6. Types of Machine Learning Algorithms

A. Supervised Learning:

Supervised Learning as the name suggests work under supervision and for its learning, the machine is trained with the data that is labeled. In supervised learning, a model can estimate with the help of a known dataset like labeled data.. Labeled data is the data that is already known to the target data. For example, when an image is shown and told that it is a cat it is labeled data. When I show you without exact detail that is unlabeled data. In supervised learning, the training dataset contains only labeled data but not unlabeled data. The objective of supervised learning is to learn a function that can vulgarize well on unseen data. The name ‘supervised Learning’ suggests that the learner is provided with the required labeled data [11].

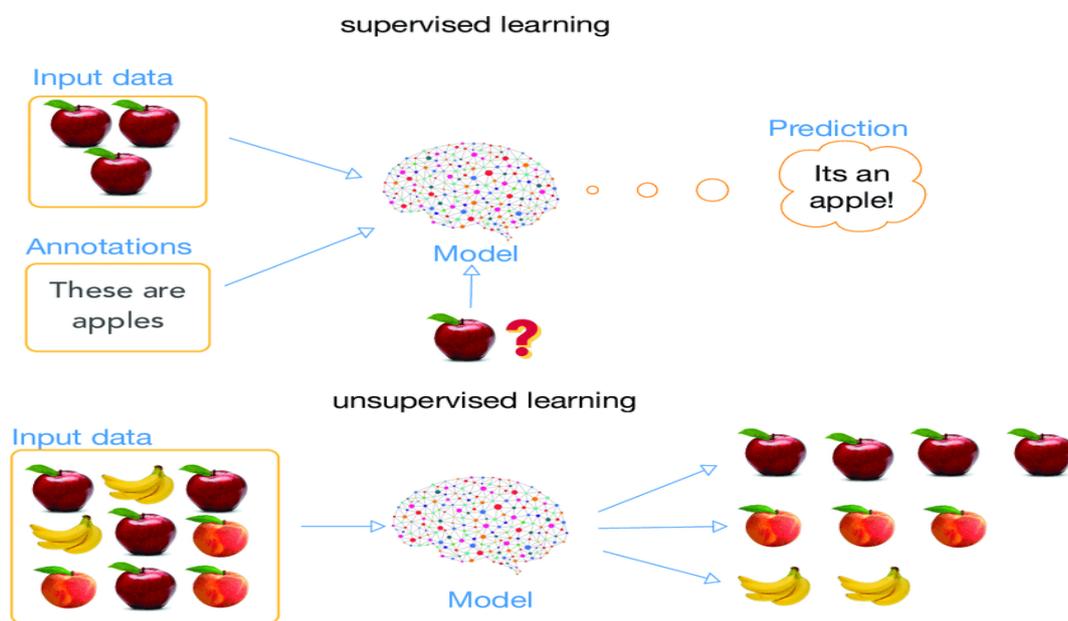


Fig.7. Supervised Learning and unsupervised learning

We have images that are labeled as roses. These are input to the machine which analyses and learns the association of these images with its labels based on its features like shape, size, color. Now a new image is fed to the machine without any label and with the help of past data, the machine can predict accurately and tell that it is a rose.

Supervised learning again two types:

Classification and Regression:

In Classification, the output variable is categorical such as yes or no / True or false, etc.

Regression: The relationship between two or more variables. Where a change in the variable is associated with another. The output of regression is real or continuous. For example, salary based on the work experience [11].

Example: Will I get a salary or not? - Classification

How much salary will I get? – Regression

B. Unsupervised Learning: In unsupervised learning, the algorithm is trained using unlabeled data and the machine tries to identify patterns and give a response. In this method, only some sample data are offered to the system as observations without any labeled data. Unsupervised learning find the regular patterns from training data set. There is no external trainer for the system to locate the pattern of the model and it is the own responsibility of the learner to find out the necessary actions. In supervised learning, the training dataset is purely labeled and in unsupervised learning, dataset is unlabeled.[12].

Unsupervised learning is further divided into two types

Clustering: The method of partitioning the data objects into clusters. The objects with in the cluster that are similar amongst them and objects which are outside of the cluster that are dissimilar to the objects belonging to another cluster.

Association: Association is one of the unsupervised learning Which identifies common co-occurrence elements

Data Mining contains several applications such as D-marts Market Analysis, Accident prone areas

C.Semi-Supervised Learning:

Semi-supervised learning is a combination of supervised and unsupervised learning. Semi-supervised learning attempts to use unlabeled data as well as labeled data to improve the classification performance. Unlabeled is cheap. Labeled data is expensive. Semi-supervised learning is the technique which discovers a better classifier from both labeled and unlabeled data. Semi-supervised learning technique can provide high performance of classification by utilizing data which is unlabeled. This process can be used to accommodate to a variety of scenario by identifying as opposed to specifying a relationship between labeled and unlabeled data from training data. Semi-supervised Learning can yield an improvement when unlabeled data can reconstruct the optimal classification boundary. These are semi-supervised learning models such as self-training, mixture models, graph-based methods, co-training and Multiview learning. The success of semisupervised learning depends completely on some underlying expectations.[13].

Table1.Combination of labeled and unlabeled data

Gender	Height	Weight
Male	165	75
Female	152	70
Male	156	69
	176	77
	146	72
	149	62

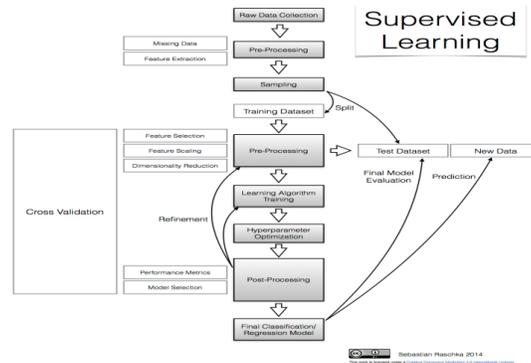


Fig.9. The Process of Supervised Machine Learning

This paper presents a idea behind the unsupervised and supervised learning models and their pattern classification analyzed and that are applied to the higher education scenario. Classification plays a important role in machine-based learning algorithms and the present study we found that, though the error back-propagation learning algorithm which works based on mathematical formula. Error back -propagation learning provided by the supervised learning model is very accurate and efficiency for several non-linear real-time problems, Kohonen self-organizing feature (KSOM) of unsupervised learning model, which provides an efficient solution and classification [16]. Designing classification networks from the given patterns which form of learning from observation. Such observation can declare a new class or assign a new class to an existing class. This classification provides new theories, techniques and knowledge that is embedded in the input patterns. The learning behavior of the neural network model improves the classification properties.

V. Case study on Disease Risk prediction

Machine learning is one of the high-speed areas of computer science (Srivastava *et al.*, 2014), with different applications, In Machine learning technique significant patterns in data will be automatically deducted with special machine learning tools, indeed gives those programs the ability to learn and adapt. In information technology the machine learning has become one of the foundations and, with that, a reasonably central, though generally hidden, part of our life. For the technological progress machine learning is a necessary ingredient for intelligent data analysis with the increasing amount of data available. There are several applications for Machine Learning (ML), being one of the most important data mining (Bustamante, Rodríguez, &Esenarro, 2019). When trying to establish the relationships between multiple characters with availability of large amount data, people tends to make mistakes while analyzing. With the appropriate learning algorithms, a different idea can be derived with data mining and machine learning hand in hand. Due to the evolution of nano technology there is lot of progress in data mining and machine learning which given a curiosity to obtain the results with identifying hidden patterns in the data. By using machine learning methods this paper addresses the automated identification of intrauterine growth restriction (IUGR) types. The research presents a comparison of un supervised and supervised learning which covers hybrid and single classification and clustering as well. Supervised learning techniques included bagging with Naïve Bayes, k-nearest neighbours (KNN), C4.5 and SMO as base classifiers, random forest as a variant of bagging with a decision tree as a base classifier, boosting with Naïve Bayes, SMO, KNN and C4.5 as base classifiers, and voting by all single classifiers using majority as a combination rule, and five single classification strategies as well: KNN, C4.5, Naïve Bayes, random tree and sequential minimal optimization algorithm for training support vector machines. Unsupervised learning included k-means and expectation-maximization algorithms [17]. In Medicine, there are many diseases and diagnoses where identification of their subtypes affects medical

treatment. Many research papers concern cancer diagnosis, appropriate feature selection techniques [18-20], and its classification based on gene expression [21]. A big challenge is an accurate classification of medical imaging and sound recordings [22] and [23]. Moreover, in many cases, the classification process is performed on labeled and unlabeled data [24].

Intrauterine growth restriction (IUGR) is a fetal growth disorder that is associated with fetal hypoxia and increased perinatal mortality. IUGR may be a significant risk factor for the development of many cardiovascular, metabolic, and pulmonologic diseases in adult life ([25]–[27]). It is a challenging problem for obstetricians, neonatologists and pediatricians, as the diagnosis is based on non-consistent definitions. This paper aims to identify an appropriate classification technique as applied to the problem of intrauterine growth restriction types. Even though classification methods have been studied extensively over the past few years ([28]), no exact solution has been discovered. Moreover, the authors usually focus on one group of machine learning techniques: supervised or unsupervised, without comparisons between the groups. This research not only constitutes an independent contribution to the relevant literature but also attempts to find a successful way to perform accurate classification of IUGR type. In the field of data mining supervised machine learning algorithms have been a dominant method. In predicting a disease using health data had recently shown a potential area of application for these methods. This particular study aims to recognize the key trends among the different supervised machine learning algorithms types and their performance including the usage for disease risk prediction [29]. This study provides the extensive indication of the relative performance of different variants of supervised machine learning algorithms for predicting the disease. Researchers can use this important information for their studies in order to select the appropriate supervised machine learning algorithms.

The scope of this research is mainly on the performance analysis of predicting the disease tactics using different variants of supervised machine learning algorithms. In recent years from the data science research community the disease prediction and medical informatics have grown in attracting the attention. This is mainly due to the wide adaptation of computer-based technology into the health sector in diverse forms (e.g., administrative data and electronic health records) and subsequent obtainability of large health databases for researchers.

VI.CONCLUSION

Data mining has importance in finding the patterns, predicting, knowledge discovery, etc in various business domains. Data mining techniques and algorithms such as cataloging, clustering, etc, helps in identifying the patterns in order to decide upcoming trends for businesses to raise and also provide many benefits in the area of airlines, banks, insurance. Classification of medical datasets is regarded as a challenging task, requiring extremely high accuracy. Therefore, researches on finding the most appropriate methods for precise classification are conducted. Multiple classifiers constitute one of the most important advances in machine learning in recent years. The identification process of IUGR pattern (symmetrical or asymmetrical) is an important medical problem to solve, as symmetric and asymmetric growth restriction may have different influences on growth and development in childhood. Moreover, asymmetric IUGR infants are more likely to have major anomalies than symmetric IUGR infants or infants appropriate for gestational age.

There are numerous applications for Machine Learning (ML) with data mining being one of the most important. A big challenge is an accurate classification of medical imaging and sound recordings. Moreover, in many cases, the classification process is performed on labeled and unlabeled data. The scope of this research is mainly on the performance analysis of disease prediction methods using different variants of supervised machine learning algorithms. Many research papers concern cancer diagnosis, appropriate feature selection

techniques and its classification based on gene expression. There has been substantial progress in data mining and machine learning as a result of development of nanotechnology, which generated curiosity in finding the hidden patterns in the data to obtain results.

References

1. Jiawei Han and Micheline Kamber (2006), **Data Mining Concepts and Techniques**, published by Morgan Kauffman, 2nd ed.
2. Jérôme Euzenat Pavel Shvaiko, Ontology Matching.
3. Zhao, Kaidi and Liu, Bing, Tirpark, Thomas M. and Weimin, Xiao, **A Visual Data Mining Framework for Convenient Identification of Useful Knowledge**, ICDM'05 Proceedings of the Fifth IEEE International Conference Data Mining, vol.-1, no.-1, pp.- 530-537, Dec 2005
4. **Crisp-DM 1.0 Step by step Data Mining guide** from <http://www.crisp-dm.org/CRISPWP-0800.pdf>.
5. R. Andrews, J. Diederich, A. B. Tickle, **A survey and critique of techniques for extracting rules from trained artificial neural networks**, Knowledge-Based Systems, vol.- 8, no.-6, pp.-378-389, 1995.
6. Venkatadri. M and Lokanatha C. Reddy ,**A comparative study on decision tree classification algorithm in datamining** , International Journal Of Computer Applications In Engineering ,Technology And Sciences(IJCAETS), Vol.- 2 ,no.- 2 , pp. 24- 29 , Sept 2010.
7. Lior Rokach and Oded Maimon, **Data Mining with Decision Trees: Theory and Applications (Series in Machine Perception and Artificial Intelligence)**, ISBN:981-2771-719, World Scientific Publishing Company, ,2008.
8. Customer Successes in your industry from http://www.spss.com/success/?source=homepage&hpzone=nav_bar.
9. <https://www.allbusiness.com/Technology/computer-software-data-management/633425-1.html>, last retrieved on 15th Aug 2010.
10. <http://www.kdnuggets.com/>.
11. R. Asadi, N. Mustapha, N. Sulaiman, and N. Shiri, “New Supervised Multi Layer Feed Forward Neural Network Model to Accelerate Classification with High Accuracy”, European Journal of Scientific Research, Vol. 33, No. 1, 2009, pp.163-178
12. R. Sathya and A. Abraham, “Unsupervised Control Paradigm for Performance Evaluation”, International Journal of Computer Application, Vol 44, No. 20, pp. 27-31, 2012.
13. https://www.researchgate.net/journal/22786856_International_Journal_of_Emerging_Trends_Technology_in_Computer_Science
14. <https://minds.wisconsin.edu/bitstream/handle/1793/60444/TR1530.pdf?sequence=1>

15. Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", ISBN 978-1-59904-252, Hershey, New York, 2007.
16. R. Sathya, Annamma Abraham, "Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification", (*IJARAI International Journal of Advanced Research in Artificial Intelligence*, Vol. 2, No. 2, 2013).
17. Agnieszka Wosiak, Agata Zamecznik, Katarzyna Niewiadomska-Jarosik," Supervised and Unsupervised Machine Learning for Improved Identification of Intrauterine Growth Restriction Types", Proceedings of the Federated Conference on Computer Science and Information Systems pp. 323–329, DOI: 10.15439/2016F515, ACSIS, Vol. 8. ISSN 2300-5963.
18. Paja W.: "Medical diagnosis support and accuracy improvement by application of total scoring from feature selection approach", Proceedings of the 2015 Federated Conference on Computer Science and Information Systems (FEDCSIS 2015), Annals of Computer Science and Information Systems, eds. M. Ganzha and L. Maciaszek and M. Paprzycki, IEEE, 2015, pp. 281–286, DOI: 10.15439/2015F361
19. Gong, P. and Cheng, Y.-H. and Wang, X.-S.: "Benign or Malignant Classification of Lung Nodules Based on Semantic Attributes", *Acta Electronica Sinica*, 2015, vol. 43, no. 12, pp. 2476–2483.
20. Pérez, N. and Guevara, M.A. and Silva, A. and Ramos, I. and Loureiro, J.: "Improving the performance of machine learning classifiers for Breast Cancer diagnosis based on feature selection", Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, Annals of Computer Science and Information Systems, eds. M. Ganzha, L. Maciaszek, M. Paprzycki, IEEE, 2014, vol. 2, pp. 209–217, DOI: 10.15439/2014F249
21. Hijazi, H. and Chan, Ch.: "A Classification Framework Applied to Cancer Gene Expression Profiles" *Journal of Healthcare Engineering*, vol. 4, no. 2, pp. 255–283, 2013, DOI: 10.1260/2040-2295.4.2.255
22. Sun, S., Wang, H., Jiang, Z., Fang, Y., Tao, T.: "Segmentation-based heart sound feature extraction combined with classifier models for a VSD diagnosis system", *Expert Systems with Applications*, 41(4), 2014, pp. 1769–1780, DOI: 10.1016/j.eswa.2013.08.076
23. Montejo, L. D., Jia, J., Kim, H. K., Netz, U. J., Blaschke, S., Muller, G. A., Hielscher, A. H.: "Computer-aided diagnosis of rheumatoid arthritis with optical tomography, Part 2: image classification", *Journal of biomedical optics*, 2013, vol. 18(7), pp. 076002–076002, DOI: 10.1117/1.JBO.18.7.076002.
24. Stamatis, K. and Nikos, F. and Sotiris, K. and Kyriakos S.: "A Semisupervised Cascade Classification Algorithm", *Applied Computational Intelligence and Soft Computing*, Article ID 5919717, 14 pages, 2016, DOI: 10.1155/2016/5919717.
25. Mahajan, S.D. and Singh, S. and Shah, P. and Gupta, N. and Kochupillai, N.: "Effect of maternal malnutrition and anemia on the endocrine regulation of fetal growth", *Endocrine research*, 2004, vol. 30(2), pp. 189–203, DOI: 10.1081/ERC-200027380
26. Baker, D.J.: "Maternal nutrition, fetal nutrition, and disease in later life", *Nutrition*, 1997, vol.13, pp. 807–813, DOI: 10.1016/S0899- 9007(97)00193-7

27. Mahajan, S.D. and Aalinkeel, R. and Singh, S. and Shah, P. And Gupta, N. and Kochupillai, N.: "Endocrine regulation in asymmetric intrauterine fetal growth retardation", *Journal of Maternal- Fetal and Neonatal Medicine*, 2006, vol. 19(10), pp. 615–623, DOI:10.1080/14767050600799901
28. Jeetha Lakshmi, P. S. and Saravan Kumar S. and Suresh A.: "A Novel Hybrid Medical Diagnosis System Based on Genetic Data Adaptation Decision Tree and Clustering", *ARNP Journal of Engineering and Applied Sciences*, vol. 10, no. 16, 2015, pp. 7293–7299
29. Shahadat Uddin, Arif Khan, Md Ekramul Hossain and Mohammad Ali Moni, "Comparing different supervised machine learning algorithms for disease prediction", Uddin et al. *BMC Medical Informatics and Decision Making* (2019) 19:281 <https://doi.org/10.1186/s12911-019-1004-8>.
30. Osisanwo F.Y., Akinsola J.E.T., Awodele O., et al, "Supervised Machine Learning Algorithms: Classification and Comparison", *International Journal of Computer Trends and Technology (IJCTT)* – Page No.128-138, Volume 48 Number 3 June 2017.