An Useful Survey on Supervised Machine Learning Algorithms: Comparisons and Classifications

B.Abhishek
School of Computer Science and Engineering,
Vellore Institute of Technology
Chennai, India
abhishek.b2020@vitstudent.ac.in

Amit Kumar Tyagi
School of Computer Science and Engineering,
Vellore Institute of Technology
Chennai, India
amitkumar.tyagi@vit.ac.in

***Abstract*—The look for methodologies that can make inferences from externally supplied data develop broad hypotheses that are subsequently used to create forecasts concerning future events is known as supervised machine learning (SML). One among the most popular common jobs performed utilizing expert system is supervised categorization. This study examine machine learning (ML) classification strategies, compares supervised learning algorithms, and determines foremost efficient classification algorithm based on the data set, number of instances, and variables (features). ML with the Waikato Environment for Knowledge Analysis (WEKA) tool, 7 different machine learning algorithms were considered: Decision Table, Random Forest (RF), Nave Bayes (NB), Support Vector Machine (SVM), Neural Networks (Perceptron), JRip, and Decision Tree (J48). The methods were implemented using the Diabetes sample data, which includes 786 instances with eight factors that are independent and there is just 1 dependent variable for the study. Time it takes to make a design and be concise (accuracy) are factors on the one end, and the kappa statistic and Mean Absolute Error (MAE) are factors on the other. For supervised predictive machine learning to work, Machine Learning algorithms must be accurate, and error-free.**

Keywords*— *Machine Learning, Mean Absolute Error, WEKA, Deep Learning

I. INTRODUCTION

These days, ML is one of the most popular fields in computer science, with a wide range of applications. It is the technique of discovering meaningful patterns in data automatically. Machine learning technologies are designed to enable algorithms to accustom. [19].
Machine Learning has emerged to become one of the cornerstones as a result of advancements in information technology, a very fundamental, if often overlooked, aspect of our lives. With an ever generation of information, smart data analysis is expected to become ever more common as a critical component of technological advancement.

Machine Learning (ML) has a variety of applications, the most important data mining is one of them. When conducting research or seeking to draw connections between various characteristics, people are more susceptible to making errors. [9].

Machine learning and Data mining are considered as pairs that may be used to extract a variety of insights using the right learning methods. Because to the advancement of smart and nano - technology, data mining and machine learning have advanced significantly, piqueing interest in finding underlying patterns to derive value. While combining machine learning, information theory, statistics and computers has resulted in a strong mathematical basis and a set of immensely effective tools. The algorithms are categorized in to a typology which depends on projected result of the algorithm. Supervised learning is used to create the function that maps inputs to expected output. Machine learning algorithms have become increasingly sophisticated as a result of unprecedented data generation. This has necessitated the use of a variety of supervised and unsupervised machine methodologies for learning. The purpose of classification tasks is to persuade the technology to understand how to use a categorization system that we've devised [21], supervised learning used frequently.

The accessibility buried within Big Data is perfectly suited for machine learning. Because It is based on data and evolves at a machine level, ML ensures extracting significance from large and diverse data sources by removing less reliance on individual tracks. Machine learning is well adapted to the challenges of dealing with several data sources, a wide range of variables, and vast amounts of data, and ML thrives on expanding data sets. Extra data fed into a machine learning framework, the better it can be taught and the better the insights it can produce. ML is brilliant at finding and displaying hidden patterns in data because it is free of the constraints of individual level reasoning and study [15].

The classification problem is a common supervised learning challenge in which the learner must look at several input-output samples that transforms learning process, divide a vector into one of several classes. The process of learning a series of rules from instances in order to build a classifier that can generalise to new situations is known as learning by deduction (examples in a training set). Figure 1 shows a way for applying supervised machine learning to a true problem.

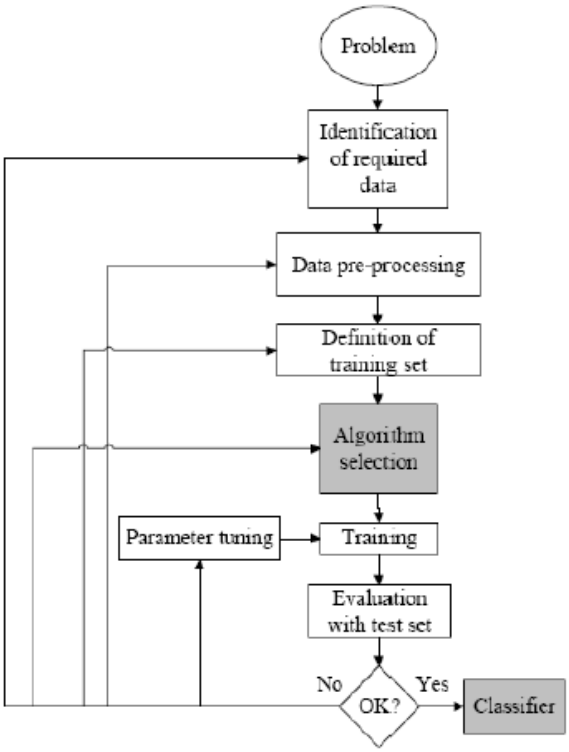

Fig.1 Supervised Machine Learning Mechanisms

This study concentrated on distinguishing machine learning algorithms and determining efficient approach with the highest precision and accuracy. In addition to testing the efficacy, on big and small datasets to classify them accurately and to provide an insight into the development of supervised ML models.

Remainder of work is divided onto the following sections: Section 2 comprises a literature review that discusses the categorization of various supervised learning algorithms; the procedure is presented in part 3, the findings are shown in section 4, and the inference and future research recommendations are presented in section 5.

II. LITERATURE SURVEY

A. Classification of Supervised Learning Algorithms

Given below are the types of SML algorithms which deals with classification, according to [21] are detailed below. As per the assertion, linear classifier achieves its goal on producing a judgment on the basis of a linear value features' combinations.

1) Linear Classifiers: Using linear (hyperplane) decision boundaries, linear classification models divide input vectors into classes [6]. In ML, purpose of classification is to organise things with similar feature values are clubbed together into groups [23]. As it is the quickest classifier and where speed is an issue, then linear classifier is used[21]. The convergence rate among data set variables and on the other hand, it is influenced by the margin. The margin is a parameter that determines how easily a dataset can be separated linearly, and to solve a classification problem in a simple manner. [18].

2) Logistic regression: A supervised learning classification approach for predicting the probability of a target variable. Location of the class boundary, as well as probabilities of class change with distance from the boundary are commonly mentioned in a specialized approach which is called as logistic regression. Probabilistic claims set logistic regression apart from other classification algorithms. It generates more exact, detailed forecasts and may be customised in a number of ways; nevertheless, such exact predictions may turn out to be erroneous. Ordinary Least Squares (OLS) and logistic regression are two methods for predicting outcomes. The outcome of logistic regression prediction, on the other hand, is a binary result [13]. In logistic regression, the linear interpolation approach is applied. [11].

3) Naive Bayesian (NB) Networks: NB is one of the basic networks and is a combination of directed acyclic graphs with only parent node and having many childrens' (representing identified nodes), using the premise that child nodes are self-contained with reference to their parents [7]. As a result, calculating probability distributions is the basis of the Naive Bayes model (independence). [14]. Bayes classifiers are generally outperformed by other, more sophisticated algorithms for learning (such as ANNs).

4) Multi-layer Perceptron: Rather of dealing with a non-convex, unconstrained minimization problem as is the case with traditional neural network training[21], the network's weights are calculated using a quadratic programming problem with linear restriction. Perceptron is [17] used in other algorithms also. It is mostly used to learn from a set of training examples by continuously executing it over them until it discovers an always correct prediction vector. The labels on the test set are then predicted using this prediction rule [9].

5) Support Vector Machines (SVMs): SVMs have recently been popular in supervised machine learning approaches [24]. Multilayer perceptron neural networks and SVM models have many commonalities. The idea of "margin" defined as distance between two classes on either side of a hyperplane, underpins SVMs. It has been demonstrated that decreasing the predicted generalisation error by widening the margin and determining the separation between the separating hyperplane and the occurrences on either side [9].

6) K-means: For handling the well-known clustering problem, K-means is considered to be the easiest unsupervised learning algorithms, according to [2] and [22]. The approach uses a simple and uncomplicated method for classifying a given sample based on preset no. of clusters (assuming k clusters). When labelled data is not available, the K-Means technique is used [1]. A generic method for transforming poor estimates into highly precise prediction rules.

7) Decision Trees (DT): It arranges instances on the basis of their feature values. Each node of the tree represents a feature in an instance to be classified, and each branch indicates a value that the

node can adopt. This is a prediction model used in decision tree learning that converts observations about an object into judgments about its goal value. It's utilised in machine learning and data mining. Classification trees or regression trees are two terms for similar tree models[20].After DT have been pruned using a validation set, post-pruning procedures are often utilised in decision tree classifiers to evaluate their performance. Any node in the sorted training instances can be eliminated and given the most common class [9].

8) Neural Networks (NN): [2] Several NN that can do multiple tasks such as regression and/or classification in one go, despite the fact that most networks can only perform one at a time. This network is a single output variable in the overwhelming majority of cases. The input and activation functions of an Artificial Neural Network, as well as the network architecture and the weight of each input link, are all critical components (ANN). The ANN's behaviour is governed by the present values of the weights since the first two features are constant. The net's weights are then considerably adjusted to get the net's output values closer to the anticipated output values. Algorithms may be used to teach a network in a variety of ways [12].

9) Bayesian Network: A graphical depiction of a collection of variables' probability relationships (features). The most well-known statistical learning approaches are Bayesian networks [9]. This has a drawback of not being suited for datasets with many features [4]. Prior understanding of a Bayesian network's topology, also known as domain knowledge, which could form/node in the following manner:

1. Designating as a root, i.e. one with no children.
2. Defining to be a leaf, i.e., one without children.
3. Designating as the direct cause or effect.
4. Declare that it is not directly connected to one another.
5. Determining the independence of two nodes based on a set of criteria.
6. Declare that a node in the ordering happens before another node by supplying partial node ordering.
7. Creating a comprehensive node sorting strategy.

A. Features of Machine Learning (ML) Algorithms

ML approaches that are supervised can be used in a variety of fields. [18], [25] include a number of ML application-oriented studies.

When dealing with multi-dimensions and continuous data, neural networks and SVMs perform substantially good. While dealing discrete/categorical characteristics, logic-based systems, in other hand, tends to project in a superior manner. To achieve optimal prediction accuracy, neural network models and SVMs require a sample in larger size, whereas NB may just require a small dataset.

The fact that k-NN is particularly sensitive to irrelevant information is well-known; this property can be explained by the algorithm's design. Furthermore, non-essential elements may render neural network training useless, if not impossible. When it comes to problems that need diagonal partitioning, most decision tree methods fall short. The instance space is divided orthogonally to one variable's axis and parallel to the other axes. As a result, after partitioning, All of the regions that arise are hyperrectangles.

Naive Bayes (NB) utilises very minimal storage capacity throughout the RAM required to retain the prior and conditional probabilities in both training stage as well as in classification stage which is absolute minimum. The fundamental kNN method requires a lot of storage during the training phase, and the execution space is at least as big as the training region. Final classifier, on the other hand, is generally a very condensed description of the data, execution space is frequently significantly lower than training space for all non-lazy learners. In addition, whereas rule

algorithms cannot be utilised as incremental learners, Naive Bayes and the kNN can. Missing values are not taken into account while calculating probabilities, as a result, in Naive Bayes, they have no impact on the final choice. Neural networks and kNN, on the other hand, function only with full records.

Finally, the operational profiles of Decision Trees and NB vary; while one is quite precise, while the other is not, and vice versa. The operational profiles of SVM and ANN are likewise comparable. Over all datasets, no single learning system can consistently outperform others. Varied data sets of different sorts of variables and quantities of occurrences influence the type of method which performs exceptionally good. According to the no-free-lunch hypothesis, on all data sets, no single learning system will outshine another. [10]. The comparison of several learning algorithms is shown in Table 1.

Table 1: Learning algorithm contrasts (**** - best performance, * - worst performance)[9]

	Decision Trees	Neural Networks	Naïve Bayes	kNN	SVM	Rule-learners
Accuracy in general	**	***	*	**	****	**
Speed of learning with respect to number of attributes and the number of instances	***	*	****	****	*	**
Speed of classification	****	****	****	*	****	****
Tolerance to missing values	***	*	****	*	**	**
Tolerance to irrelevant Attributes	***	*	**	**	****	**
Tolerance to redundant attributes	**	**	*	**	***	**
Tolerance to highly interdependent attributes (e.g. parity problems)	**	***	*	*	***	**
Dealing with discrete/binary/continuous Attributes	****	***(not discrete)	***(not continuous)	***(not directly discrete)	**(not discrete)	***(not directly continuous)
Tolerance to noise	**	**	***	*	**	*
Dealing with danger of overfitting	**	*	***	***	**	**
Attempts for incremental Learning	**	***	****	****	**	*
Explanation ability/transparency of knowledge/classifications	****	*	****	**	*	****
Model parameter handling	***	*	****	***	*	***

III. METHODOLOGY

The National Institute of Diabetes and Digestive and Kidney Diseases provided the relevant data for the study. This data was chosen for its correctness, as well as the fact that it has been anonymized (de-identified), ensuring confidentiality. There are eight attributes, plus one class, for a total of nine. The following are the numeric values for all attributes:

1. How many times you've been pregnant
2. In an oral glucose tolerance test, plasma glucose concentration after 2 hours

3. Blood pressure in the diastole (mm Hg)
4. The thickness of the skin folds on the triceps (mm)
5. Insulin serum (mu U/ml) after 2 hours
6. BMI (weight in kilogrammes divided by height in metres)
7. Pedigree function in diabetes
8. Your age (years)
9. Variable in the class

Class	Value Number of Instances	Converted Value (attribute)
0	500	NO
1	268	YES

Table 2: Class Distribution: (1- "tested positive for diabetes") and (0- "tested negative for diabetes")

Table 2 displays the overall number of cases included in this study, with 500 of them testing positive for diabetes and 268 testing negative.

WEKA 3.7.13 was used to perform a comparative examination of various supervised machine learning methods. The values 1s in the class distribution (class variable) were changed to YES, indicating that they had been tested for DIABETES POSITIVELY, and the values 0s to NO, indicating that they had been tested for DIABETES NEGATIVELY. Random Forest, Decision Table, Neural Networks (Perceptron), SVM, Nave Bayes, JRip and Decision Tree were the seven classification methods utilised in this study (J48).

This study was able to anticipate precision and accuracy assured in terms of distinct machine learning algorithms by setting parameters by using 2 alternative sets of no. of occurrences.

IV. RESULTS

Various machine learning methods were classified and compared using WEKA. The results are shown in Table 3 with nine qualities and factors taken into account.

Algorithm	Time (Sec)	Correctly Classified (%)	Incorrectly Classified (%)	Test Mode	Attributes	No of instancing	Kappa statistic	MAE	Precision of YES	Precision of NO	Classification
Decision Table	0.23	72.3958	27.6042	10-fold cross-validation	9	768	0.3752	0.341	0.619	0.771	Rules
Random Forest	0.55	74.7396	25.2604	10-fold cross-validation	9	768	0.4313	0.3105	0.653	0.791	Trees
Naïve Bayes	0.03	76.3021	23.6979	10-fold cross-validation	9	768	0.4664	0.2841	0.678	0.802	Bayes
SVM	0.09	77.3438	22.6563	10-fold cross-validation	9	768	0.4682	0.2266	0.740	0.785	Functions

Neural Networks	0.81	75.1302	24.8698	10-fold cross-validation	9	768	0.4445	0.2938	0.653	0.799	Functions
JRip	0.19	74.4792	25.5208	10-fold cross-validation	9	768	0.4171	0.3461	0.659	0.780	Rules
Decision Tree (J48)	0.14	73.8281	26.1719	10-fold cross-validation	9	768	0.4164	0.3158	0.632	0.790	Tree

Table 3 : Shows a comparison of different categorization algorithms using a larger data set and more attributes.

The TIME it takes to develop the model is referred to as TIME. The Mean Absolute Error (MAE) is a metric that indicates how accurate a prognosis or prediction is in terms of the actual outcome. The Kappa Statistic is a statistic for comparing observed and expected accuracy (Random Chance): YES indicates that you have been diagnosed with diabetes. NO indicates that a diabetes test was negative. Further, Table 4 displays the classification results using six attributes and a contrasting the various machine learning techniques and settings examined.

Table 4 : shows a comparison of different categorization algorithms using a smaller data set and fewer attributes

Algorithm	**Time**	**Correctly Classified %**	**Incorrectly Classified %**	**Test Mode**	**Attributes**	**No of instance**	**Kappa statistic**	**MAE**	**Precision of YES**	**Precision of NO**	**Classification**
Decision Table	0.09	67.9688	32.0313	10-fold cross-validation	6	384	0.3748	0.3101	0.581	0.734	Rules
Random Forest	0.42	71.875	28.125	10-fold cross-validation	6	384	0.3917	0.3438	0.639	0.761	Trees
Naïve Bayes	0.01	70.5729	29.4271	10-fold cross-validation	6	364	0.352	0.3297	0.633	0.739	Bayes
SVM	0.04	72.9167	27.0833	10-fold cross-validation	6	384	0.3837	0.2708	0.711	0.735	Functions
Neural Networks (Perceptron)	0.17	**59**	**41**	10-fold cross-validation	6	384	**0.1156**	**0.4035**	0.444	0.672	Functions
JRip	0.01	**64**	**36**	10-fold cross-validation	6	384	**0.2278**	**0.4179**	0.514	0.714	Rules
Decision Tree (J48)	0.03	**64 %**	**36**	10-fold cross-validation	6	384	**0.1822**	**0.4165**	0.519	0.685	Tree

The TIME it takes to develop the model is referred to as TIME. The MAE is a metric that actually gives a picture that how close a forecast or prediction is to the actual result.

Kappa Statistic is a measure which actually compares the observed accuracy to that of anticipated accuracy (Random Chance). YES indicates that the patient has been diagnosed with diabetes. NO indicates that a diabetes test was negative.

Table 5 : **Smaller Dataset 384**

Smaller Dataset 384		
Algorithm	**Precision of YES (Positive Diabetes)**	**Precision of NO (Negative Diabetes)**
SVM	0.711	0.735
Random Forest	0.639	0.761
Naïve Bayes	0.633	0.739
Decision Table	0.581	0.734
Decision Tree (J48)	0.519	0.685
JRip	0.514	0.714
Neural Networks (Perceptron)	0.444	0.672

Table 6: **Large Data Set 768**

Large Data Set 768		
Algorithm	**Precision of YES (Positive Diabetes)**	**Precision of NO (Negative Diabetes**
SVM	0.74	0.785
Naïve Bayes	0.678	0.802
JRip	0.659	0.78
Random Forest	0.653	0.791
Neural Networks (Perceptron)	0.653	0.799
Decision Tree (J48)	0.632	0.79
Decision Table	0.619	0.771

Tables 5 and 6 Using smaller and larger data sets, illustrate how different algorithms order the precision of positive and negative diabetes.

Now below, Tables 7 and 8 show rankings i.e. correctly classified and erroneously classified data sets based on the time it took to develop the model using various techniques for smaller and larger data sets, respectively.

Table 7 : **Smaller Dataset 384**

Smaller Dataset 384			
Algorithm	**Time**	**Correctly Classified**	**Incorrectly Classified**
SVM	0.04 sec	72.92%	27.08%
Random Forest	0.42 sec	71.88%	28.13%
Naïve Bayes	0.01 sec	70.57%	29.43%
Decision Table	0.09 sec	67.97%	32.03%
JRip	0.01 sec	**64%**	**36%**
Decision Tree (J48)	0.03 sec	**64%**	**36%**
Neural Networks (Perceptron)	0.17 sec	**59%**	**41%**

Table 8: **Large Data Set 768**

Large Data Set 768			
Algorithm	**Time**	**Correctly Classified**	**Incorrectly Classified**
SVM	0.09 sec	77.34%	22.66%
Naïve Bayes	0.03 sec	76.30%	23.70%
Neural Networks (Perceptron)	0.81 sec	75.13%	24.87%
Random Forest	0.55 sec	74.74%	25.26%
JRip	0.19 sec	74.48%	25.52%
Decision Tree (J48)	0.14 sec	73.83%	26.17%
Decision Table	0.23 sec	72.40%	27.60%

Table 9: A Descriptive Analysis of Different Dataset Attributes

Attribute number	Mean	Standard Deviation
1	3.8	3.4
2	120.9	32.0
3	69.1	19.4
4	20.5	16.0
5	79.8	115.2
6	32.0	7.9
7	0.5	0.3
8	33.2	11.8

B. Discussions

Table 3 shows a comparison of findings of 768 samples and of 9 features. It may be seen that if compared to MAE, algorithms are having a higher Kappa statistic. In addition, occurrences that have been correctly classified outnumber those that have been mistakenly classified. This indicates that predictive analysis is more reliable with larger data sets. As demonstrated in table 3, SVM and NB require a large sample size to attain maximum prediction accuracy, whereas DT and Decision Table have the least precision.

Table 4 gives a comparative results for 384 instances and 6 attributes. When compared to MAE, Neural Networks, JRip, and J48 all have lower Kappa statistics, which do not give precision and accuracy. It shows that JRip, J48 and Neural Networks perform well with smaller datasets have a far lower percentage of correctly categorized occurrences than wrongly classified ones. SVM and RF, on the other hand, demonstrate high accuracy and precision with fewer data sets. In comparison to JRip and Decision Tree, Decision Table took longer to build the model. As a result, saving time does not imply accuracy. If Kappa Statistic is smaller than Mean Absolute Error (MAE), Algorithm will not show any precision and accuracy. As a result, algorithm with such features cannot be applied for that data set since precision and accuracy will be lacking.

Table 6 compares precision for larger and data sets that are smaller, with SVM indicating algorithm with the best prediction. Table 5 also demonstrates that SVM is the most precise algorithm. Data sets that are smaller.

Tales 7 and 8 compare the percentage of successfully identified and erroneously categorised data given the time it took to build the model for smaller and larger datasets,. Table 7 reveals that Naive Bayes and JRip are the algorithms with the fastest time to build, however JRip has a lower percentage of correctly classified samples, indicating the time it took to create as a model is not synonymous with accuracy. Similarly, at a time of 0.04 seconds, SVM has the best level of accuracy. When compared to Table 8, the third correctly categorised algorithm was neural networks (perceptron). This indicates that a Neural Network works better with a large dataset than it does with a small one. In addition, the findings show that Decision Table struggles with huge datasets.

Table 9 depicts mean and standard deviation for qualities studied, with concentration of Plasma glucose (attribute 2) having the highest mean and Diabetes pedigree function (attribute 7) having the lowest mean, showing a significant effect on a small data set.

In the last, in [26-41] authors have recommended authors to read these research efforts, to know more information about the role of AI, Computer Vision, or Machine learning techniques respect to these sensitive areas/ useful applications like healthcare, agriculture, etc. We hope that readers/ researchers will find suitable problem for themselves from these research work.

V. CONCLUSION AND FUTURE SCOPE

The parameters for ML classification must be fine-tuned thoroughly, and the data collection must have a large number of occurrences. It takes time to build the algorithm's model, but it also

requires accuracy and precise classification. As a result, the best learning approach for one set of data could not work for another set of data with logically different features. In machine learning classification, the most essential question is not whether one learning algorithm is better than another, but rather under what conditions one technique may surpass another on a specific application job. In this direction, meta-learning is moving, with researchers attempting to discover functions that link datasets and algorithm performance [12]. Meta-learning does this by representing the features of learning issues with a collection of traits called as meta-qualities, with the objective of discovering links between these traits and the effectiveness of learning algorithms. The number of examples, the proportion of categorical characteristics, the fraction of missing data, and the entropy of class are all used to construct learning tasks.

[3] given a dataset with a long list of data and statistical measurements. After gaining a deeper grasp of each method's strengths and limitations, the option of combining To solve an issue, two or more algorithms should be investigated. The idea is to balance off the flaws of one method with the advantages of another. SVM, NB, and RF machine learning algorithms provide great precision and accuracy regardless of the number of characteristics and data examples. Machine learning algorithms must be precise, accurate, and have a low error rate in order to be used for supervised predictive machine learning..

Large data quantities should be processed in a distributed processing environment, according to this research. This will result in a higher level of correlation between the variables, resulting in a better model output.

REFERENCES

[1] Alex S.& Vishwanathan, S.V.N. (2008). *Introduction to Machine Learning.* Published by the press syndicate of the University of Cambridge, Cambridge, United Kingdom. Cambridge University Press 2008. ISBN: 0-521-82583-0.

[2] Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, England. 1995. Oxford University Press, Inc. New York, NY, USA ©1995 ISBN:0198538642

[3] Brazdil P., Soares C. &da Costa, J. (2003). Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results.*Machine Learning*Volume 50, Issue 3,2003.Copyright ©Kluwer Academic Publishers. Manufactured in The Netherlands, doi:10.1023/A:1021713901879pp. 251–277.

[4] Cheng, J., Greiner, R., Kelly, J., Bell, D.& Liu, W. (2002). Learning Bayesian networks from data: An information-theory based approach. *Artificial Intelligence* Volume 137. Copyright © 2002. Published by Elsevier Science B.V. All rights reserved pp. 43 – 90.

[5] Domingos, P. & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning* Volume 29, pp. 103–130 Copyright © 1997 Kluwer Academic Publishers. Manufactured in The Netherlands

[6] Elder, J. (n.d). Introduction to Machine Learning and Pattern Recognition. Available at LASSONDE University EECS Department York

[7] Good, I.J. (1951). Probability and the Weighing of Evidence, *Philosophy* Volume 26, Issue 97, 1951. Published by Charles Griffin and Company, London 1950.Copyright © The Royal Institute of Philosophy 1951,pp. 163-164.doi: https://doi.org/10.1017/S0031819100026863.

[8] Hormozi, H., Hormozi, E. & Nohooji, H. R. (2012). The Classification of the Applicable Machine Learning Methods in Robot Manipulators. *International Journal of Machine Learning and Computing (IJMLC),* Vol. 2, No. 5, 2012 doi: 10.7763/IJMLC.2012.V2.189pp. 560 – 563.

[9] Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica* 31 (2007). Pp. 249 – 268. Retrieved from IJS

[10] Lemnaru C. (2012). Strategies for dealing with Real World Classification Problems, (Unpublished PhD thesis) Faculty of Computer Science and Automation, Universitatea Technica, Din Cluj-Napoca.
[11] Logistic Regression pp. 223 – 237. Available at: https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf
[12] Neocleous C. & Schizas C. (2002). Artificial Neural Network Learning: A Comparative Review. In: Vlahavas I.P., Spyropoulos C.D. (eds)*Methods and Applications of Artificial Intelligence*. Hellenic Conference on Artificial IntelligenceSETN 2002. Lecture Notes in Computer Science, Volume 2308. Springer, Berlin, Heidelberg, doi: 10.1007/3-540-46014-4_27 pp. 300-313.
[13] Newsom, I. (2015). Data Analysis II: Logistic Regression.
[14] Nilsson, N.J. (1965). Learning machines. New York: McGraw-Hill.Published in: Journal of IEEE Transactions on Information Theory Volume 12 Issue 3, 1966. doi: 10.1109/TIT.1966.1053912 pp. 407 – 407.
[15] Pradeep, K. R. & Naveen, N. C. (2017). A Collective Study of Machine Learning (ML)Algorithms with Big Data Analytics (BDA) for Healthcare Analytics (HcA). *International Journal of Computer Trends and Technology (IJCTT)* – Volume 47 Number 3, 2017. ISSN: 2231-2803, doi: 10.14445/22312803/IJCTT-V47P121, pp 149 – 155. Available from IJCTT website: http://www.ijcttjournal.org/2017/Volume47/number-3/IJCTT-V47P121.pdf
[16] Rob Schapire (n.d) Machine Learning Algorithms for Classifrication.
[17] Rosenblatt, F. (1962), Principles of Neurodynamics. Spartan, New York.
[18] Setiono R. and Loew, W. K. (2000), FERNN: An algorithm for fast extraction of rules from neural networks, Applied Intelligence.
[19] Shai Shalev-Shwartz and Shai Ben-David (2014). Understanding Machine Learning From Theory to Algorithms.
[20] T. Hastie, R. Tibshirani, J. H. Friedman (2001) ― The elements of statistical learning,‖ Data mining, inference, and prediction, 2001, New York: Springer Verlag.
[21] Taiwo, O. A. (2010). Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), ISBN: 978-953-307-034-6, *InTech,* University of Portsmouth United Kingdom. Pp 3 – 31.
[22] Tapas Kanungo, D. M. (2002). A local search approximation algorithm for k-means clustering. Proceedings of the eighteenth annual symposium on Computational geometry. Barcelona, Spain: ACM Press.
[23] Timothy Jason Shepard, P. J. (1998). Decision Fusion Using a Multi-Linear Classifier. In Proceedings of the International Conference on Multisource-Multisensor Information Fusion.
[24] Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. (2nd ed.). Springer Verlag. Pp. 1 – 20. Retrieved from website: https://www.andrew.cmu.edu/user/kk3n/simplicity/vapnik2000.pdf
[25] Witten, I. H. & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques* (2nd ed.), ISBN: 0-12-088407-0, Morgan Kaufmann Publishers, San Francisco, CA, U.S.A. © 2005 Elsevier Inc.Retrieved
[26] B. Gudeti, S. Mishra, S. Malik, T. F. Fernandez, A. K. Tyagi and S. Kumari, "A Novel Approach to Predict Chronic Kidney Disease using Machine Learning Algorithms," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2020, pp. 1630-1635, doi: 10.1109/ICECA49313.2020.9297392.
[27] Gillala Rekha, V. Krishna Reddy, and Amit Kumar Tyagi, “KDOS - Kernel Density based Over Sampling - A Solution to Skewed Class Distribution”, 2020, Journal of Information Assurance and Security (JIAS), Vol. 15 Issue 2, p44-52. 9p.
[28] Gillala Rekha, Amit Kumar Tyagi, and V Krishna Reddy, “Solving Class Imbalance Problem Using Bagging, Boosting Techniques, with and without Noise Filter Method”, International Journal of Hybrid Intelligent Systems, vol. 15, no. 2, pp. 67-76, 2019.
[29] Nair M.M., Kumari S., Tyagi A.K., Sravanthi K. (2021) Deep Learning for Medical Image Recognition: Open Issues and a Way to Forward. In: Goyal D., Gupta A.K., Piuri

V., Ganzha M., Paprzycki M. (eds) Proceedings of the Second International Conference on Information Management and Machine Intelligence. Lecture Notes in Networks and Systems, vol 166. Springer, Singapore. https://doi.org/10.1007/978-981-15-9689-6_38

[30] Akshara Pramod, Harsh Sankar Naicker, Amit Kumar Tyagi, "Machine Learning and Deep Learning: Open Issues and Future Research Directions for Next Ten Years", Book: Computational Analysis and Understanding of Deep Learning for Medical Care: Principles, Methods, and Applications, 2020, Wiley Scrivener, 2020.

[31] Tyagi, Amit Kumar and G, Rekha, Machine Learning with Big Data (March 20, 2019). Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur - India, February 26-28, 2019

[32] Amit Kumar Tyagi, Poonam Chahal, "Artificial Intelligence and Machine Learning Algorithms", Book: Challenges and Applications for Implementing Machine Learning in Computer Vision, IGI Global, 2020.DOI: 10.4018/978-1-7998-0182-5.ch008

[33] Kumari S., Vani V., Malik S., Tyagi A.K., Reddy S. (2021) Analysis of Text Mining Tools in Disease Prediction. In: Abraham A., Hanne T., Castillo O., Gandhi N., Nogueira Rios T., Hong TP. (eds) Hybrid Intelligent Systems. HIS 2020. Advances in Intelligent Systems and Computing, vol 1375. Springer, Cham. https://doi.org/10.1007/978-3-030-73050-5_55

[34] Varsha R., Nair S.M., Tyagi A.K., Aswathy S.U., RadhaKrishnan R. (2021) The Future with Advanced Analytics: A Sequential Analysis of the Disruptive Technology's Scope. In: Abraham A., Hanne T., Castillo O., Gandhi N., Nogueira Rios T., Hong TP. (eds) Hybrid Intelligent Systems. HIS 2020. Advances in Intelligent Systems and Computing, vol 1375. Springer, Cham. https://doi.org/10.1007/978-3-030-73050-5_56

[35] Shamila M, Vinuthna, K. & Tyagi, Amit. (2019). A Review on Several Critical Issues and Challenges in IoT based e-Healthcare System. 1036-1043. 10.1109/ICCS45141.2019.9065831.

[36] Tyagi, Amit Kumar; Nair, Meghna Manoj; Niladhuri, Sreenath; Abraham, Ajith, "Security, Privacy Research issues in Various Computing Platforms: A Survey and the Road Ahead", Journal of Information Assurance & Security . 2020, Vol. 15 Issue 1, p1-16. 16p.

[37] Tyagi, Amit Kumar; Nair, Meghna Manoj "Internet of Everything (IoE) and Internet of Things (IoTs): Threat Analyses, Possible Opportunities for Future, Vol. 15 Issue 4, 2020.

[38] Tyagi A.K., Fernandez T.F., Mishra S., Kumari S. (2021) Intelligent Automation Systems at the Core of Industry 4.0. In: Abraham A., Piuri V., Gandhi N., Siarry P., Kaklauskas A., Madureira A. (eds) Intelligent Systems Design and Applications. ISDA 2020. Advances in Intelligent Systems and Computing, vol 1351. Springer, Cham. https://doi.org/10.1007/978-3-030-71187-0_1

[39] Shasvi Mishra, Amit Kumar Tyagi, "The Role of Machine Learning Techniques in Internet of Things Based Cloud Applications", AI-IoT book, Springer, 2021.

[40] L. Kanuru, A. K. Tyagi, A. S. U, T. F. Fernandez, N. Sreenath and S. Mishra, "Prediction of Pesticides and Fertilizers using Machine Learning and Internet of Things," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-6, doi: 10.1109/ICCCI50826.2021.9402536.

[41] R, Varsha et al. 'Deep Learning Based Blockchain Solution for Preserving Privacy in Future Vehicles'. 1 Jan. 2020: 223 – 236.