MACHINE LEARNING: APPLICATIONS, TECHNIQUES AND CURRENT SCENARIO

1st Bhavana Shah
Sal Institute of Technology And Engineering Research
Ahmedabad, India

2nd Dr. Vidhi Khanduja
Sal Institute of Technology And Engineering Research
Ahmedabad, India

***Abstract*—Machine Learning (ML) used in day-to-day life. ML can use data and use it for self-learning. It is widely used in many fields like finance, agriculture, education, and security, etc. This paper discusses the potential of utilizing machine learning technologies in various sector.ML categorized in mainly four Learning process. Supervised learning which contains labelled data where unsupervised learning includes unlabelled data. Semi-supervised learning is a combination of supervised and unsupervised learning. The reinforcement learning algorithm is learned by receiving feedback on the effect of modifying some parameters. We can study in detail with different techniques and different methods and also check which algorithm is more accurate in less runtime. This paper summarizes some application of machine learning such as prediction, disease detection, fraud detection and more. This paper helps in reducing the research gap for ML applications.**

***Keywords*—Machine Learning, Detection, Prediction, SVM, Naïve Bayes, KNN, ANN.**

I. INTRODUCTION

Machine Learning (ML) can access data and use it for self-learning. The main motto is to allow the computers to learn automatically without human intercession or collaboration and adjust actions according to it.

Fig.1 Basic ML Model

Fig.1 shows the steps involved in ML process. ML Model consists mainly of four steps: environment, learning, repository and execute [1]. Environment means to acquire outside information. Quality of information affect directly on learning to understand whether easy or clutter. Learning means to get the knowledge from information and that knowledge is added to the repository which stores many general principles that guide the implementation activities. Thomas H. Davenport writes in the Wall Street Journal, "Humans can typically create one or two good models a week (while) machine learning can create thousands of models a week"[2].

Machine Learning Techniques:

Fig. 2 shows the Machine learning classification techniques is divided into following category [4, 5]:

A. Supervised Learning,
B. Unsupervised Learning,
C. Reinforcement Learning.
D. Semi-supervised Learning

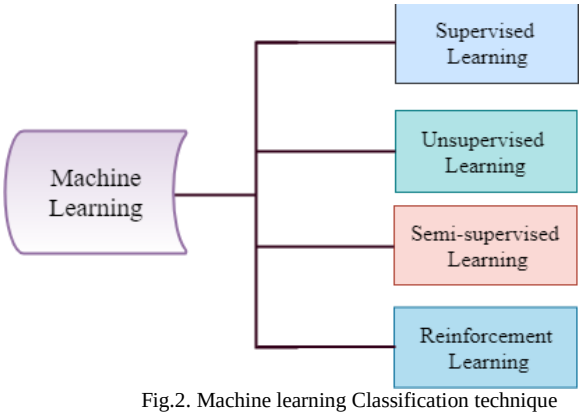

Fig.2. Machine learning Classification technique

A. *Supervised learning*

It recognizes a labelled data set and is divided into two types: classification and regression [3]. Supervised learning is widely used in different-different applications, like speech recognition, spam detection and object recognition [6]. Supervised learning, if applied with proper data, the error rate can be reduced tens to the minimum error rate bound [3]. Table 1 describes the tasks performed using supervised learning, the algorithms used for performing the tasks and application domains.

Table1. Supervised Learning Details

Category	Tasks	Algorithms	Applications Examples
Supervised Learning	Classification	Neural networks, decision tree, SVM	Intrusion/fault/anomaly detection
	Regression	Logistic regression, SVR, Gaussian process for regression	Throughput prediction, channel parameter regression

B. *Unsupervised Learning*

Unsupervised Learning means we have a collection of unlabelled data, we wish to analysis and discover patterns within it. Unsupervised learning has two important examples; dimension reduction and clustering [7]. Classic clustering algorithms include k means, hierarchical clustering, spectrum clustering, fuzzy clustering and the Dirichlet process [3, 4]. Table 2 displays algorithms used to segment text topics, propose items and find data outliers.

Table 2. Unsupervised Learning Details

Category	Tasks	Algorithms	Applications Examples
Unsupervised learning	Clustering	k means, spectrum clustering	Congestion control [5], hierarchical routing[5]
	Dimension reduction	Manifold learning, local linear embedding(LLE), isometric mapping(ISO MAP)	Data aggregation[5]

C. *Reinforcement Learning*

The algorithm learns by receiving feedback on the effect of modifying some parameters, e.g. the power and modulation [6]. In this machine is trained to take exact decisions based on the business requirement to maximize efficiency. It is used in gaming and navigation and in general, to address applications such as robotics, finance (investment decisions), inventory management, where the goal is to learn a policy, i.e., a mapping between states of the environment into actions to be performed, while directly interacting with the environment [6,7]. Table 3 shows the algorithm task and technique used in various applications.

Table 3. Reinforcement Learning Details

Category	Tasks	Algorithms	Applications Examples
Reinforcement learning	Policy learning	Q learning	Resource management, routing

D. *Semi-supervised learning*

It is a combination of supervised and unsupervised learning. Self-training is the oldest form of semi-supervised learning [8]. Example of semi-supervised learning is recognizing a person's face on a webcam. It consists of the methods such as classification, regression, and prediction [5] semi-supervised learning techniques can be organized in four classes: i) methods based on generative models ii) methods based on the assumption that the decision boundary should lie in a low-density region iii) graph-based methods iv) two-step methods (first an unsupervised learning step to change the data representation or build a new kernel; then a supervised learning step based on the new representation or kernel) [6].

II. Application

ML has various applications in diverse fields**:**

Agriculture

In India, agriculture has become an important source of economic development. The agriculture industries started exploring new methods to grow food production because of increasing population and changes in weather. This makes researchers find new well-organized and valuable technologies for high productivity. Agriculturalist can collect data by using precision agriculture. Precision agriculture is technology, which provides advanced techniques to improve farm output. By utilizing Precision agriculture it is possible to make economic growth in agriculture and it is also used in many applications like pest detection in plants, weed detection, yield production of crops and plant disease detection, etc. A farmer uses pesticides to control pest, prevent diseases and to increase crop yield. The diseases in the crop are creating a problem of low production and economic losses to farmers and agricultural industries. Therefore identification of disease and its severity based as become necessary. Table 4 shows different-different techniques are used detect the plant diseases and accuracy of each.

Table 4.ML Classification Technique for Agriculture

ML Classification Technique	Culture	No. of Diseases	Result
SVM Classifier	Oil palm [10]	2 diseases	97% accuracy for Chimaera and 95% accuracy for Anthracnose disease.
	Potato [11]	2 diseases	Accuracy 90%.
	Tea [12]	3 diseases	Accuracy 93%.
	Soybean [13]	3 diseases	Accuracy is approximately 90%
ANN Classifier	Pomegranate[14]	4 diseases	Accuracy around 90%
	Groundnut [15]	4 diseases	Accuracy 97.41%.
KNN Classifier	Cotton [16]	1 disease	Accuracy 82.5%
Fuzzy Classifier	Wheat [17]	1 disease	Disease detection accuracy 88% and recognition of disease type accuracy56%
CNN Classifier	14 crops [18]	26 diseases	Accuracy 99.35%.
	Soybean [19]	3 diseases	Accuracy 99.32%.
	25 plants [20]	58 diseases	Accuracy 99.53%

Healthcare

Healthcare field facing more problems and it is becoming more expensive. Several ML techniques used to rectify them, ML techniques use for prediction of various diseases like thyroid disease, cardiovascular disease, heart disease, Hepatitis Disease, cancer disease, and diabetic disease. From the earlier study, identified that naive Bayes provides 86% of accuracy for the diagnosis of heart disease. SVM gives 96.40% of accuracy for the breast cancer diagnosis, and CART provides 79% of accuracy for a tasting of diabetic disease [4]. Table 5 display the various techniques to detect the particular disease with its accuracy, also mention the dataset collect from which place.

Table 5 .ML Classification Technique for Healthcare

ML Technique used	Disease	Dataset	Accuracy
J48	Heart Disease [21]	UCI	84.35%
SVM			85.03%
Naïve Bayes	Heart Disease[22]	Diabetic Research Institute in Chennai	86.41%
Decision Tree	Breast Cancer [23]	Swami Vivekananda Diagnostic Centre Hospital	97%
DT+SVM	Breast Cancer [24]	UCI Machine Learning Repository	91%
SVM	Thyroid Disease[25]	UCI	98.62%
SVM	Diabetic Disease[26]	UCI	78%

Finance

Machine learning and artificial intelligence in financial industry [28] holds a lot of scopes, these include credit unions, banks, credit card companies, insurance companies, accountancy companies, consumer finance companies, stock brokerages, investment funds, and individual fund managers. [27] Fig.3 shows ML techniques are used for the finance segments.

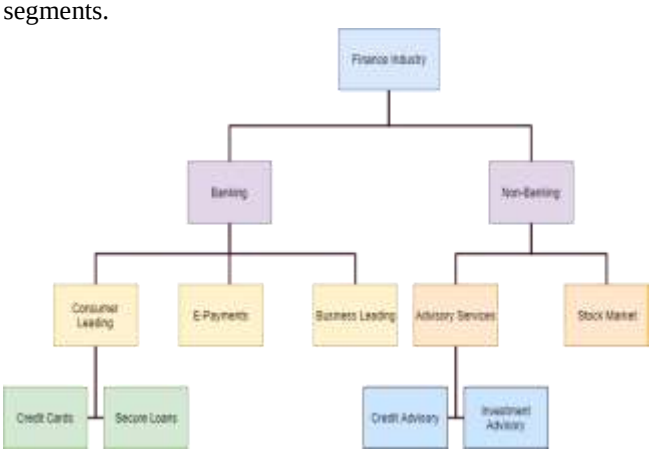

Fig 3.Financial segments

A. Diagnostic Analytics

It is used for knowing the root cause of a phenomenon, such as a project success or a failure. Diagnostic analytics includes all forms of financial risk assessment. A practical implementation is a tool "Halo" by Pricewaterhouse Cooper. Halo works on accounting journals and ledgers [29]. Its ability to flag higher-risk transactions. This is not an online real-time solution since the effort is only to diagnose and point out the financial risks. Accounting is about old transaction records, and that is what a tool like a halo tries to extract information from a client's systems and discover to bring forth high-risk transactions with their patterns and trends. It also gives out information on process inefficiencies in the accounting process, like eliminating duplicate accounting journal entries, etc. This is useful for a large financial organization because of the high volume of transactions that happen during a particular period in various departments [27].

B. Predictive Analytics

It is used to show the result of an event in the future, such as the success or failure of a financial transaction, financial budget overrun, or a budget slippage for an organization such as a bank or a company. There are different parameters in machine learning model code takes into account for analysis before deriving the conclusion. Purpose is to provide a sense to the top management direction that the financial side of the business. It does not solve the problem, but it merely tells or acts a warning signal of failure that deviates from the plan for future related to finances. After applying predictive analytics, we get lots of fields of finance that come into play, such as cost accountancy, managerial economics, and decision support systems. predictive costing reports are based on past cost accountancy data.it helps large financial businesses stay on track with their cost budgets and make sure the right flags are raised if the real cost is predicting to increase as a result of prediction by a machine learning model. In management accounting or economics, as it is known in the academic circles, the focus on what-if scenarios [27]. ML models based on past data and applying them based on principles of management accounting.

C. *Prescriptive Analytics*

The greatest value obtains by constructing the prediction model based on predictive analytics and is advised for future actions. it uses ML to build a future plan based on past failures in financial transactions in similar situations, such as failed financial transactions, and it gives out specific activities that need to carry out in order for the financial transaction to success. In the financial sector, it is now truly able to do course corrections even before the failures happen. The prescriptive part starts when there is an accurate prediction made; then the actions that need to take even before this happens come after that [27].

Education

ML used in education have much focus lately For predicting student's performance, the widely used model is instance-based learning, Naïve Bayes, Decision Tree, Artificial Neural Network, Support Vector Machine, Classification Tree [30], and clustering. Various data used to analyse the student's performance, such as prelude of the earlier semester, pre-university, examination mark, and demographic attribute. Generalized Linear Model (GLM) is the extended framework of linear regression. The idea computing a hyperplane to cut the loss function [31]. Support vector machine is widely used for data mining and classification to predict membership of a data. It is based on the geometrical interpretation. The algorithm searches for optimal separating surface, such as hyperplane [32]. The decision tree is one of the most commonly used techniques in predicting student's performance. Data mining in education field known as Educational Data Mining (EDM). It happens because of the increase in educational resources and data that can be explored to learn how a student learned. Researchers also investigated the factors that influence learning outcomes [33] Table 6 show the accuracy result for the Generalized Linear (GL), Deep Learning (DL), and Decision Tree (DT) with runtime.

Table 6. Experimental Accuracy Result for Education

Model	Accuracy	Run Time
Generalized Linear[34]	66.6%	1 s
Deep Learning[34]	67.6 %	5 s
Decision Tree[34]	60.6 %	2 s

Security

Using unauthorized APs in the company, government and military facilities, there is a high possibility of being subjected to different viruses and hacking attacks. It requires to detect unauthorized APs for the protection of information. Various methods of research are in progress, addressing various aspects of the issue [36]. Dataset was created using Round Trip Time (RTT) values. The data set organized to apply the machine learning algorithm to give the result, and then the results obtained are compared, to show which algorithm is more accurate [35]. Table 7 shows the various ML methods are used for security purpose with its accuracy.

Table 7. Experimental Accuracy Result for Security

Algorithms	SVM	C4.5	KNN	MLP
True Positive	40	92.9	92.9	84.5
False Positive	0	9.1	8.5	8.4
Total Correctness	70	92.9	84.1	88

III. Current Scenario

Lot of research has been done exploring basic ML techniques. Research is now shifted to advanced algorithms. We now present the state-of-art of latest trends in ML. Advanced topics in machine learning including: Linear Modelling, Nonlinear Dimension Reduction, Maximum Entropy, Exponential Family Models, Conditional Random Fields, Graphical Models, Structured Support Vector Machines, Feature Selection, Kernel Selection, Meta-Learning, Multi-Task Learning, Graph-Based Semi-Supervised Learning, Approximate Inference, Clustering, and Boosting.

Table 8. Advance Machine Learning Techniques

Classification		Algorithm	USE
Single classifiers	Supervised Learning	K-nearest neighbor [37]	• nonparametric technique • frequently used approach for classification of samples[4] • k-NN is called instance based learning • calculate distance measure from N number of training samples
		Support vector machines (SVM) [37]	• designed for binary classification • provides a user specified parameter called penalty factor • simple and prominent process.[4] • applying a kernel function they can embed data into a higher dimensional space, in which data points can be linearly separated[6] • mainly used for classification and regression problems[43]
		Artificial neural networks [37]	• also called as backpropagation neural networks • pattern recognition[9]
		Decision trees [37]	• sequence of decisions • current decision helps to make the subsequent decision • represented in a tree structure, top node =root node [4] • well-known program for constructing decision trees is CART (Classification and Regressing Tree)
		• Classification Tree [37]	• decision tree with a range of discrete (symbolic) class labels is called a classification tree • Intrusion/fault/anomaly detection[3]
		• Regression tree [37]	• Decision tree with a range of continuous (numeric) values is called a regression tree. • Throughput prediction, channel parameter regression[3]
	Unsupervised Learning	Self-organizing maps [37]	• trained by an unsupervised competitive learning algorithm • reduce the dimension of data visualization • Data aggregation[3]
		Clustering [6]	• process of grouping data • intracluster similarity is high • inter-cluster similarity is low • Congestion control, hierarchical routing[3]
		• k-means [6]	• algorithm may terminate at a local optimum partition • initialization of the parameters can be done using k-means

			• Congestion control, hierarchical routing[3]
		• Gaussian mixture model [6]	• linear superposition of Gaussian • distributions
	Statistics base learning	Naïve bayes networks [37]	• use a probabilistic graph model called Naïve Baysian Networks • represented by a directed acyclic graph (DAG) • identify the class membership probabilities based on given class label[4]
	Other techniques	Genetic algorithms (GA) [37]	• implement the natural selection and evolution • concept comes from the ‘‘adaptive survival in natural organisms” • low performing programs are replaced by genetic recombination’s of high performing programs
		Fuzzy logic [37]	• evolved from Fuzzy set theory • considers the set membership values for reasoning • values range between 0 and 1 • used in engineering applications[4]
Hybrid classifiers		Cluster + Single methods [37,38]	• K-mean+ MLP(multilayer perceptron) [39] • SOM + MLP [40]
		Cascaded hybrid methods [37,38]	• Association based case reduction+CBR[41]
		Integrated-based hybrid methods [37,38]	• CRB ensemble+SVM[42]
Ensemble classifiers		Boosting [37]	• Boosting tries to reduce bias. • Bootstrapped Aggregation (Bagging)[5] • Bagging solve over-fitting problem and increase bias
		Bagging [37]	

IV. CONCLUSION

This study shows that ml techniques are crucial in different fields. ML has provides the different techniques and various method for getting correct result. It also confirms that ml methods are exclusively used in the diagnosis and predictions of various types of diesel. ML techniques for prediction of various diseases like heart disease, cancer, diabetic, thyroid disease etc. using SVM, Naïve Bayes, DT. For heart disease we get 86.41% correct result by using Naïve Bayes method. In finance sector ml provide tool "halo" by pricewaterhouse cooper. For predicting student’s performance, implementing generalized linear method is 66.6% right with run time 1s. Security purpose it requires to detect unauthorized aps for the protection of information. C 4.5 gives the 92.9% of total correctness. At last discuss advance machine learning classifiers to solve past problems and get more accurate result with less runtime.

REFERENCES

[1] Wang, H., Ma, C. and Zhou, L., 2009, December. A brief review of machine learning and its application. In 2009 international conference on information engineering and computer science (pp. 1-4). IEEE.

[2] Bhardwaj, R., Nambiar, A.R. and Dutta, D., 2017, July. A study of machine learning in healthcare. In 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC) (Vol. 2, pp. 236-241). IEEE.

[3] E, H., Liang, L., Li, G.Y., Kim, J., Lu, L. and Wu, M., 2017. Machine learning for vehicular networks. arXiv preprint arXiv:1712.07143.

[4] Shailaja, K., Seetharamulu, B. and Jabbar, M.A., 2018, March. Machine Learning in Healthcare: A Review. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 910-914). IEEE.

[5] Nithya, B. and Ilango, V., 2017, June. Predictive analytics in health care using machine learning tools and techniques. In 2017 International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 492-499). IEEE.

[6] Musumeci, F., Rottondi, C., Nag, A., Macaluso, I., Zibar, D., Ruffini, M. and Tornatore, M., 2018. An overview on application of machine learning techniques in optical networks. IEEE Communications Surveys & Tutorials, 21(2), pp.1383-1408.

[7] Kanchan, B.D. and Kishor, M.M., 2016, December. Study of machine learning algorithms for special disease prediction using principal of component analysis. In 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC) (pp. 5-10). IEEE.

[8] Chapelle, O., Scholkopf, B. and Zien, A., 2009. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. IEEE Transactions on Neural Networks, 20(3), pp.542-542.

[9] Shruthi, U., Nagaveni, V. and Raghavendra, B.K., 2019, March. A Review on Machine Learning Classification Techniques for Plant Disease Detection. In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS) (pp. 281-284). IEEE.

[10] Ahmad Nor Ikhwan Masazhar and Mahanijah Md Kamal, “Digital Image Processing Technique for Palm Oil Leaf Disease Detection using Multiclass SVM”, IEEE 4th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Malaysia 2017, pp. 1-6.

[11] Monzurul Islam, Anh Dinh and Khan Wahid, “Detection of potato Diseases Using Image Segmentation and Multiclass Support Vector Machine”, IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Canada 2017, pp. 1-4

[12] Nithesh Agarwal, Jyothi Singhai and Dheeraj K. Agarwal, “Grape Leaf Disease Detection and Classification Using Multi- Class Support Vector Machine”, proceeding of IEEE International conference on Recent Innovations in Signal Processing and Embedded Systems (RISE), Bhopal 2017, pp. 238-244

[13] Md. Selim Hossain, Rokeya Mumtahana Mou, Mohammed Mahedi Hasan, Sajib Chakraborty and M. Abdur Razzak, “Recognition and Detection of Tea Leaf’s Diseases Using Support Vector Machine”, IEEE 14th International Colloquium on Signal Processing & its Applications (CSPA), Malaysia 2018, pp. 150-154.

[14] Mrunmayee Dhakate and Ingole A. B. , “Diagnosis of Pomegranate Plant Diseases using Neural Network”, IEEE 5th National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Patna 2015.

[15] Ramakrishnan M. and Sahaya Anselin Nisha A., "Groundnut Leaf Disease Detection and Classification by using Back Probagation Algorithm". IEEE International Conference on Communications and Signal Processing (ICCSP), Melmaruvathur 2015, pp. 0964 – 0968.

[16] Aditya Parikh, Mehul S. Raval, Chandrasinh Parmar and Sanjay Chaudhry, "Disease Detection and Severity Estimation in Cotton Plant from Unconstrained Images", IEEE International Conference on Data Science and Advanced Analytic, Canada 2016, pp. 594-601.

[17] Diptesh Majumdar, Arya Ghosh, Dipak Kumar Kole, Aruna Chakraborty and Dwijesh Dutta Majumder, "Application of Fuzzy CMeans Clustering Method to Classify Wheat Leaf Images based on the presence of rust disease", Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications, Vol. 327, 2015, pp. 277-284.

[18] Sharada P. Mohanty, David P. Hughes and Marcel Salathe, " Using Deep Learning for Image- Based Plant Disease Detection", Frontiers in Plant Science, Vol. 7, Article 1419, 2016.

[19] Serawork Wallelign, Mihai Polceanu and Cedric Buche, "Soybean Plant Disease Identification Using Convolutional Neural Network", International Florida Artificial Intelligence Research Society Conference (FLAIRS-31), Melbourne, United States 2018, pp. 146-151.

[20] Konstantinos P. Ferentinos, "Deep learning models for plant disease detection and Diagnosis", Computers and Electronics in Agriculture, Vol. 145, Elsevier 2018, pp. 311-318.

[21] Chaurasia, V. and S. Pal," Data Mining Approach to Detect Heart Disease", International Journal of Advanced Computer Science and Information Technology, vol.2, pp.56-66, 2018.

[22] Vembandasamy et al., "Heart Diseases Detection Using Naive Bayes Algorithm", vol.2, pp. 441-444, 2015.

[23] E. Venkatesan et al., "Performance analysis of decision tree algorithms for breast cancer classification", Indian Journal of Science and Technology, vol.8, pp.1-8, 2015.

[24] K. Sivakami, "Mining Big Data: Breast Cancer Prediction using DTSVM Hybrid Model", International Journal of Scientific Engineering and Applied Science vol.1, pp.418-429, 2015.

[25] M. R. Nazari Kousarrizi et al., "An Experimental Comparative Study on Thyroid Disease Diagnosis Based on Feature Subset Selection and classification", International Journal of Electrical& Computer Sciences, vol.1, pp.13-19, 2012.

[26] Kumari, V.A. and R. Chitra, "Classification of Diabetes Disease Using Support Vector Machine", International Journal of Engineering Research and Applications, vol.3, pp. 1797-1801, 2013.

[27] Mathur, P., 2019. Overview of Machine Learning in Finance. In Machine Learning Applications Using Python (pp. 259-270). Apress, Berkeley, CA

[28] Noche, Y.K.H., Guinto, M.B.P., Paulo, D.M.E.M. and Sahagun, C.M., The Views of Twelve (12) Companies in Makati City on Hiring Senior High School Graduates.

[29] Kokina, J. and Davenport, T.H., 2017. The emergence of artificial intelligence: How automation is changing auditing. Journal of Emerging Technologies in Accounting, 14(1), pp.115-122.

[30] I. Đurđević Babić, "Machine learning methods in predicting the student academic motivation," Croat. Oper. Res. Rev., vol. 8, no. 2, pp. 443–461, 2017.

[31] A. Kumar, J. Naughton, and J. M. Patel, "Learning Generalized Linear Models Over Normalized Data," Proc. 2015 ACM SIGMOD Int. Conf. Manag. Data - SIGMOD '15, pp. 1969–1984, 2015.

[32] M. Rachmadi, M. Valdés-Hernández, M. Agan, and T. Komura, "Deep Learning vs. Conventional Machine Learning: Pilot Study of WMH Segmentation in Brain MRI with Absence or Mild Vascular Pathology," J. Imaging, vol. 3, no. 4, p. 66, 2017.

[33] A. B. Urbina Nájera and J. De la Calleja Mora, "Brief Review of Educational Applications Using Data Mining and Machine learning," Rev. Electrónica Investig. Educ., vol. 19, no. 4, p. 84, 2017.

[34] Tanuar, E., Heryadi, Y., Abbas, B.S. and Gaol, F.L., 2018, September. Using Machine Learning Techniques to Earlier Predict Student's Performance. In 2018 Indonesian Association for Pattern Recognition International Conference (INAPR) (pp. 85-89). IEEE.

[35] Kim, D., Shin, D. and Shin, D., 2018, August. Unauthorized Access Point Detection Using Machine Learning Algorithms for Information Protection. In 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE) (pp. 1876-1878). IEEE.

[36] F. Awad, M. Al-Refai, and A. Al-Qerem. "Rogue access point localization using particle swarm optimization," in 8th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan. May 2017. doi: 10.1109/IACS.2017.7921985

[37] Intrusion detection by machine learning: A review Chih-Fong Tsai a , Yu-Feng Hsu b , Chia-Ying Lin c , Wei-Yang Lin d,*

[38] Machine Learning in Financial Crisis Prediction: A Survey Wei-Yang Lin, Ya-Han Hu, and Chih-Fong Tsai

[39] N.-C. Hsieh, "Hybrid mining approach in the design of credit scoring models," Expert Syst. Appl., vol. 28, pp. 655–665, 2005.

[40] J. Huysmans, B. Baesens, J. Vanthienen, and T. Van Gestel, "Failure prediction with self-organizing maps," Expert Syst. Appl., vol. 30, pp. 479–487, 2006.

[41] C.-H. Liu, L.-S. Chen, and C.-C. Hsu, "An association-based case reduction technique for case-based reasoning," Inf. Sci., vol. 178, pp. 3347– 3355, 2008.

[42] H. Li and J. Sun, "Predicting business failure using multiple case-based reasoning combined with support vector machine," Expert Syst. Appl., vol. 36, pp. 10085–10096, 2009

[43] K .P Murphy, Machine Learning: A Probabilistic Perspective, the MIT Press, 2012