PERINTIS eJournal

PERINTIS eJournal, **2020**, Vol. 10, No. 1, pp. 1-24

A SYSTEMATIC LITERATURE REVIEW ON SUPERVISED MACHINE LEARNING ALGORITHMS

Nur Amalina Diyana Suhaimi*[1,2], Hafiza Abas[1]

[1] Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Wilayah Persekutuan Kuala Lumpur, Malaysia
[2] Centre for Research and Innovation, Open University Malaysia, Petaling Jaya, Selangor Darul Ehsan, Malaysia

* Corresponding author: nuramalinadiyana@graduate.utm.my

ABSTRACT

There are many researchers and data analyst in large companies around the world applied Machine Learning (ML) in the various study. ML is a subset of Artificial Intelligence (AI) which play a significant role in analyzing the big data. In general, the Supervised Machine Learning (SML), one type of ML, generates the desired output and makes a prediction based on the trained dataset provided in the input. Various algorithms under SML including Naïve Bayes, Logistic Regression, Random Forest, J48, CART, Multi-Layer Perceptron, Support Vector Machine (SVM) which are common and famously used by researchers. There is a total of 305 studies that is being compiled initially in this paper reviews on SML classification algorithms by adopting Systematic Literature Review (SLR) method. After sorting the papers according to the selection criteria and data extraction, 61 final studies were selected. As a conclusion, SML had been mostly used in classifying spam and text and also in healthcare and medically related classification research. It is also founded that SVM and Artificial Neural Network (ANN) are the top two performing algorithms in classification. In the future work, it is recommended to expand the study to include more assessment measures of SML algorithms and unsupervised machine learning (UML).

***Keywords**: Systematic Literature Review, Supervised Machine Learning, Machine Learning, Algorithms*

1.0 INTRODUCTION

In this 21st century, Machine Learning (ML) is widely being used in the various study by researchers and data analyst in large companies around the world. With regard to the big data explosions, ML algorithm which is known as a subset of Artificial Intelligence (AI), is playing a role in performing tasks related to the big data. Some of ML applications include customer profiling, email spam filtering, credit risk analysis and fraud detection. Generally, ML can be categorized into three different types namely Supervised Machine Learning (SML), Unsupervised Machine Learning (UML) and Reinforcement Machine Learning (RML). The main difference between SML and UML is that the machine in SML is trained to learn the labeled data and then make predictions on the new test dataset provided whereby in contrast to

UML, the machine is not given any labeled, classified or categorized data. Meanwhile, the RML is a combination between both SML and UML. The machine is expected to sort and group the data according to similarities, patterns, and differences by itself [1].

In this paper, it will be focused more on SML which is commonly being used in solving classification problems because the targeted output is often to get the machine to learn and make predictions. There are several classification algorithms under the SML that can be used for example Naïve Bayes, Logistic Regression, Random Forest, J48, CART, Multi-Layer Perceptron, Support Vector Machine, and many more. Different researchers may use different algorithms in their study, and they performed a comparison of the accuracy and other comparison metrics in selecting the best predictive models. These can be seen in articles as studied by Bijamov et al. [2] who found that Neural Networks as one of SML algorithms classify the object of UXO classification better than UML when using the EMI data. This is supported by another study conducted by Randle et al. [3] which compare the performance of the SML and UML techniques in their research specifically on gaming and the results shows that the performance of SML is better than UML. In another research works by Sathya and Abraham [4] in 2013, they discover that the SML algorithms were showing better results for many non-linear real time problems compared to UML algorithms.

The main objective of this study is to gain a thorough insight into different algorithms used in SML and their categories. Another objective is to determine and compare the best performance measures of the SML algorithm. In fulfilling both research objectives, a comprehensive review of the existing research works from the year 2015 to 2019 is done by performing the SLR. Only related articles which published in the past five years research duration was selected due to the need to examine thoroughly on the recent and newest changes which are relevant in the field of the study.

Next, in Section 2 the details of the methodology used which is SLR is discussed and elaborated further. SLR is one of the useful methodologies which assist researchers in gathering all relevant publications more efficiently prior to answering the research questions and fulfilling the research works requirement apart from performing the usual traditional Literature Review (LR). This is supported by Mallet et al. [5] who highlighted that SLR deliver more advantages to researchers as LR often focus exclusively on results of other studies, meanwhile SLR results from a much broader field based on the keywords searches. Other than that, from the SLR study, it helps the researchers to be on track and useful for other researchers in future study to understand the trend in the field over the years more clearly. Thus, SLR is the best and result in a more quality review compared to the traditional LR. However, there are also challenges in applying SLR as such it requires lots of effort, patient, and time consuming. Other than that, access to many academic databases are essential and crucial for researchers to conduct the SLR. This is due to the need to complete the process of SLR comprehensively [5].

Following remaining part of this study is Section 3 which presents the details of exploration on the outcomes of the SLR work related to SML algorithms. In the final part of Section 4, the study is concluded with the limitations arises and conclusion.

2.0 METHODOLOGY

Although previously there are many SLR studies that have been published by researchers, but there were still limited, and lack of online publication found on the SLR research specifically applies on SML at the time of writing. Instead, there are many SLR published paper on ML was found, one of it was written by Dallora et al. [6] which focusing in the area of ML and microsimulation techniques application in prognosis of dementia. Meanwhile, Malhotra [7] in his SLR paper conclude that from the findings, the number of papers applying the machine learning techniques in predicting the software fault is still limited. In the recent study conducted by Carvalho et al. [8] the SLR research focuses on the ML methods applied to predictive maintenance, meanwhile Azeem et al. [9] report on the SLR and meta-analysis results of ML techniques for code smell detection. Thus, this SLR paper aims to provide an insight for future researchers' work in SML. This paper is adopting the SLR method from Kitchenham and Charters [10] and Jnr and Pa [11]. The SLR mainly focuses on the current state of research works conducted within the year range from 2015 until 2019 associated with SML. There are four processes involved in the SLR according to Jnr and Pa [11] which elaborated further in this section, begin with the research questions and motivation of the study.

2.1 Research questions (RQ) and Motivation

There are a large number of research papers discussed SML in various categories. The review intends to find out what SML algorithms have been used, how they work and evaluated. Thus, in this paper, the SLR aims to answer the following addressed research questions in Table 1:

Table 1 Research questions and motivation.

No	Research Questions	Motivation
1	What are the SML algorithms that have been used and in which categories?	Identify and classify the usage of SML algorithms based on categories
2	What are the measures in the SML algorithms that have been applied in selecting the best predictive models? What are the performance results of these algorithms?	Determine and compare the best performance measures of the SML algorithms

The next step process of SLR subsequently identifying the research questions, as shown in Figure 1 are (1) the keyword patterns for paper search process, (2) initiate inclusion and exclusion criteria, (3) the selection process or data mining and finally (4) the analysis of SLR findings to answer research questions.

Figure 1 SLR process based on Jnr and Pa (2015) [11].

2.2 The search process

With regards to the searching process, the searches took place in three digital libraries or databases available online which are IEEEXplore Digital Library, ScienceDirect Journal, and SpringerLink Journal. Only these databases were selected and accessible by the researchers due to limited time and resources available at the time of research. The searches were conducted in October and November 2019. The findings from peer-reviewed journals, published books and article from conferences were limited to the recent five years studies published between 2015 and 2019. Referring to the research questions, it can be recognized that the important key phrase was "Supervised Machine Learning" or "SML". Prior to running the key phase search, the search strings were adjusted according to the requirements of the database.

Table 2 Selected digital libraries and initial search results.

No	Digital libraries / Database	URLs	Initial Search results
1	IEEEXplore Digital Library	https://ieeexplore.ieee.org/Xplore/home.jsp	148
2	ScienceDirect Journal	https://www.sciencedirect.com/	80
3	SpringerLink Journal	https://link.springer.com/	77
Total number of papers			305

As shown in Table 2, a total of 305 papers found in the initial search done. The next step was to apply the abstract screenings and full-text screening process where the inclusion and exclusion criteria were also checked. Next is the identification and selection process of the quality papers based on quality assessment criteria and score. Finally, is the presentation of the final results of SLR.

2.3 The inclusion/exclusion criteria

There are few criteria taken into consideration during the searching process. The inclusion criteria for this study includes:

- Titles mentioned the key phrase
- Articles in English
- In the form of peer-reviewed journals, books or article from conferences
- Published between 2015 and 2019

Meanwhile, the exclusion criteria are as follows:

- Titles did not mention the key phrase
- Articles in other languages, not in English
- Not in the form of peer-reviewed journals, books or article from conferences
- Published in the year other than between 2015 and 2019
- Found to be a repeated study

2.4 The selection process (data mining)

Following that, there is a second phase of full-text screening on 137 papers where the content screening on all full-text was conducted. Only those related to "Supervised Machine Learning" or "SML" were selected. Finally, it can be concluded that a result of 76 papers was excluded from this study due to not fulfill the criteria and 61 papers were selected for further assessment.

Table 3 Abstract and title screening.

No	Digital libraries / Database	Initial search results	Not in English	Not found	Not related	Cannot be access	Secondary search	Excluded during the full-text screening	Final selected studies
1	IEEEXplore Digital Library	148	6	-	63	3	76	37	39
2	ScienceDirect Journal	80	-	2	52	1	25	13	12
3	SpringerLink Journal	77	-	-	-	41	36	26	10
Total number of papers		305	6	2	115	42	137	76	61

3.0 RESULTS AND DISCUSSION

3.1 Descriptive Analysis

A brief descriptive analysis was conducted in analyzing the SLR findings. Based on Figure 2, it shows an upward trend in the number of SML papers published from 2015 until 2019. Most of the papers (n = 21, 34.43%) were published in the year 2019, and 12 of it (57.14%) are from IEEEXplore Digital Library which all of it are published in the Conference Proceedings. Further exploration found that this is due to the number of Conference Proceedings available in IEEEXplore Digital Library are more than 8000, as compared to the number of journals available in the database is less than 400. Meanwhile, in the year 2018, there are 17 papers published (27.87%) related to "Supervised Machine Learning" or "SML" and followed by 11 paper (18.03%) published in the year 2017. There is no article (n = 0) related to SML were published by SpringerLink Journal for the year 2017 based on this study. And finally, the remaining 12 papers (19.67%) were published equally in the year 2015 and the year 2016. Table 4, 5 and 6 below presents part of the details of the final selected studies from the three digital libraries or databases.

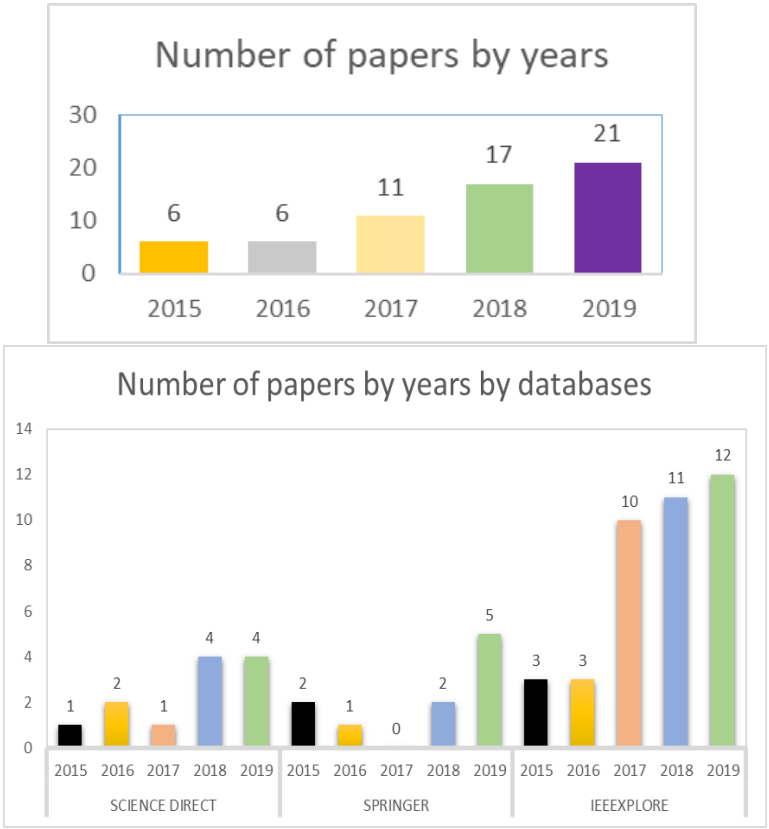

Figure 2 Number of papers by years by databases.

Table 4 Details of final papers selected from IEEEXplore Digital Library.

IEEEXplore Digital Library				
ID [Ref]	**Year**	**Focus Area**	**Title**	**Conference Proceedings**
1 [12]	2019	Sentiment Analysis	Implementation of sentiment classification of movie reviews by supervised machine learning approaches	3rd International Conference on Computing Methodologies and Communication
2 [13]	2016	Sentiment Analysis	On multi-tier sentiment analysis using supervised machine learning	IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology
3 [14]	2017	Text Classification	Performance analysis of supervised machine learning algorithms for text classification	19th International Conference on Computer and Information Technology
4 [15]	2019	Machinery	Gas turbine fault classification based on machine learning supervised techniques	14th International Conference on Electronics Computer and Computation
5 [16]	2019	Network	Network intrusion detection using supervised machine learning technique with feature selection	International Conference on Robotics, Electrical and Signal Processing Techniques
6 [17]	2019	Plantation	Plant disease classification using soft computing supervised machine learning	3rd International Conference on Electronics, Communication and Aerospace Technology
7 [18]	2018	Credit card	Supervised machine learning algorithms for credit card fraudulent transaction detection: A comparative study	IEEE International Conference on Information Reuse and Integration
8 [19]	2019	Sentiment Analysis	Sentimental analysis of book reviews using unsupervised semantic	2nd International Conference on Green

			orientation and supervised machine learning approaches	Computing and Internet of Things
9 [20]	2018	Spam SMS	SMS spam filtering using supervised machine learning algorithms	8th International Conference on Cloud Computing, Data Science and Engineering
10 [21]	2017	Modulation recognition	Supervised machine learning for signals having RRC shaped pulses	50th Asilomar Conference on Signals, Systems and Computers
11 [22]	2015	Spam Email	A survey and evaluation of supervised machine learning techniques for spam e-mail filtering	IEEE International Conference on Electrical, Computer and Communication Technologies
12 [23]	2016	Cloud Security	Feasibility of supervised machine learning for Cloud Security	International Conference on Information Science and Security
13 [24]	2017	Sentiment Analysis	Evaluating cross domain sentiment analysis using supervised machine learning techniques	Intelligent Systems Conference
14 [25]	2018	Network	Intrusion detection in network systems through hybrid supervised and unsupervised machine learning process: A case study on the ISCX dataset	1st International Conference on Data Intelligence and Security
15 [26]	2018	Business Reviews	Rumor detection in business reviews using supervised machine learning	5th International Conference on Behavioral, Economic, and Socio-Cultural Computing
16 [27]	2016	Sentiment Analysis	Predicting the sentiment of SaaS online reviews using supervised machine learning techniques	International Joint Conference on Neural Networks
17 [28]	2015	Email Classification	An empirical study on email classification using supervised machine learning in real environments	IEEE International Conference on Communications

18 [29]	2019	Text classification	Performance analysis of supervised machine learning approaches for Bengali text categorization	5th Asia-Pacific World Congress on Computer Science and Engineering
19 [30]	2018	Radar performance	Evaluating radar performance under complex electromagnetic environment using supervised machine learning methods: A case study	8th International Conference on Electronics Information and Emergency Communication
20 [31]	2019	IoT	An empirical comparison of supervised machine learning algorithms for Internet of Things data	4th International Conference on Computing Communication Control and Automation
21 [32]	2018	Arabic tweets	A supervised machine learning approach for events extraction out of Arabic tweets	5th International Conference on Social Networks Analysis, Management and Security
22 [33]	2017	Test collection	Using supervised machine learning to automatically build relevance judgments for a test collection	28th International Workshop on Database and Expert Systems Applications
23 [34]	2019	Healthcare	Monitoring system for sickle cell disease patients by using supervised machine learning	2nd Al-Sadiq International Conference on Multidisciplinary in IT and Communication Science and Applications
24 [35]	2017	Driving behavior	Modeling tactical lane-change behavior for automated vehicles: A supervised machine learning approach	5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems
25 [36]	2018	Generator insulation	Comparison of supervised machine learning techniques for PD classification in generator insulation	IEEE International Conference on Industrial and Information Systems

26 [37]	2017	Road Friction	Road friction estimation for connected vehicles using supervised machine learning	IEEE Intelligent Vehicles Symposium (IV)
27 [38]	2017	Healthcare	Fall detection using supervised machine learning algorithms: A comparative study	8th International Conference on Modelling, Identification and Control
28 [39]	2017	Scientific paper	Extend relation identification in scientific papers based on supervised machine learning	International Conference on Advanced Computer Science and Information Systems
29 [40]	2017	Sentiment Analysis	Performance analysis of supervised machine learning techniques for sentiment analysis	3rd International Conference on Sensing, Signal Processing and Security
30 [41]	2015	Software	Machine Learning Algorithms	Encyclopedia of Applied and Computational Mathematics
31 [42]	2019	Healthcare	Suitable supervised machine learning techniques for Malignant Mesothelioma diagnosis	11th Biomedical Engineering International Conference
32 [43]	2019	Network	Supervised machine learning techniques for efficient network intrusion detection	28th International Conference on Computer Communication and Networks
33 [44]	2019	Electrical	Supervising communication SoC for secure operation using machine learning	IEEE 62nd International Midwest Symposium on Circuits and Systems
34 [45]	2018	Sliding Window	Window-size impact on detection rate of wearable-sensor-based fall detection using supervised machine learning	IEEE Life Sciences Conference

35 [46]	2019	Vehicle	Supervising vehicle using pattern recognition: Detecting unusual behavior using machine learning algorithms	IEEE Region Ten Symposium
36 [47]	2018	Network	Predictive analytics of sensor data based on supervised machine learning algorithms	International Conference on Next Generation Computing and Information Systems
37 [48]	2017	Traffic	IP traffic classification in NFV: A benchmarking of supervised machine learning algorithms	IEEE Colombian Conference on Communications and Computing
38 [49]	2018	Healthcare	A comparative study of Breast Cancer diagnosis using supervised machine learning techniques	2nd International Conference on Computing Methodologies and Communication
39 [50]	2018	Water supply	A supervised energy monitoring-based machine learning approach for anomaly detection in a clean water supply system	International Conference on Cyber Security and Protection of Digital Services

Table 5 Details of final papers selected from ScienceDirect Journal.

ScienceDirect Journal				
ID [Ref]	**Year**	**Focus Area**	**Title**	**Journals / Book Chapters**
1 [51]	2019	Clinical / Medical	Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia	Computers in Biology and Medicine
2 [52]	2019	Financial	Integration of unsupervised and supervised machine learning algorithms for credit risk assessment	Expert Systems with Applications
3 [53]	2018	Sentiment Analysis	Lexicon-based approach outperforms supervised	Telematics and Informatics

			machine learning approach for Urdu Sentiment Analysis in multiple domains	
4 [54]	2019	Veterinary	Estrous detection by continuous measurements of vaginal temperature and conductivity with supervised machine learning in cattle	Theriogenology
5 [55]	2016	Social Media	Social media research: The application of supervised machine learning in organizational communication research	Computers in Human Behavior
6 [56]	2019	Biomass	Hydrogen production via biomass gasification, and modeling by supervised machine learning algorithms	International Journal of Hydrogen Energy
7 [57]	2018	Energy / Electrical	Power disaggregation of combined HVAC loads using supervised machine learning algorithms	Energy and Buildings
8 [58]	2017	Clinical / Medical	Prediction of lung cancer patient survival via supervised machine learning classification techniques	International Journal of Medical Informatics
9 [59]	2018	Food image	Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets	Computers in Biology and Medicine
10 [60]	2018	Computer vision	Recognition of CAPTCHA characters by supervised machine learning algorithms	IFAC PapersOnLine
11 [61]	2015	Archeology	A supervised machine-learning approach towards geochemical predictive modeling in archaeology	Journal of Archaeological Science

12 [62]	2016	Network security	Performance evaluation of supervised machine learning algorithms for intrusion detection	Procedia Computer Science

Table 6 Details of final papers selected from SpringerLink Journal.

SpringerLink Journal				
ID [Ref]	**Year**	**Focus Area**	**Title**	**Journals / Book Chapters**
1 [63]	2019	Neuroimaging	Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets	Brain Imaging and Behavior
2 [64]	2019	Healthcare	Supervised machine learning techniques and genetic optimization for occupational diseases risk prediction	Soft Computing
3 [65]	2019	Movie Prediction	Prediction model for Bollywood movie success: A comparative analysis of performance of supervised machine learning algorithms	The Review of Socionetwork Strategies
4 [66]	2019	Text classification	Survey on supervised machine learning techniques for automatic text classification	Artificial Intelligence Review
5 [67]	2019	Environmental Health	Using mid-infrared spectroscopy and supervised machine-learning to identify vertebrate blood meals in the malaria vector, Anopheles arabinoses	Malaria Journal
6 [68]	2018	Healthcare	Comparative motor pre-clinical assessment in Parkinson's disease using supervised machine learning approaches	Annals of Biomedical Engineering

7 [69]	2018	Stars and planetary system	Protostellar classification using supervised machine learning algorithms	Astrophysics and Space Science
8 [70]	2016	Land zone	Automatic arable land detection with supervised machine learning	Earth Science Informatics
9 [71]	2015	Healthcare	A systematic comparison and evaluation of supervised machine learning classifiers using headache dataset	Advanced Intelligent Computing Theories and Applications
10 [72]	2015	Nuclear Power Plant	Performance comparison of supervised machine learning algorithms for multiclass transient classification in a nuclear power plant	5th International Conference on Swarm, Evolutionary, and Memetic Computing

3.2 RQ1: SML algorithms based on categories

Moving forward, in answering the first research question (RQ1), the process of identifying and classifying the usage of SML algorithms based on categories was conducted by performing a comprehensive study on the full details of the 61 articles which are obtained from the three digital libraries or databases. The full details of the articles consisting of the author's name, title of the article, the title of the journal/conference proceedings, the volume of the journal, year published, page numbers, ISSN numbers, Digital Object Identifier (DOI) of the article and/or URL link of article. In SpringerLink, it specifies and mentioned the content type of the article whether it is an "Article" or a "Book Chapter" whereas, in ScienceDirect and IEEEXplore, the journals and digital library include the information of the Abstract and Keywords of an article in the full details. Following that, the focus areas of the SML algorithms used were extracted from the articles' description or extracted from the "Title" of the articles and classes into 16 different categories as summaries in Table 7.

Most of the articles (n = 10, 16.40%) have used SML in classifying spam and text in their research study, followed by healthcare and medical-related classification research (n = 9, 14.75%). There are 8 articles that focus on the network; software and computers use SML, meanwhile 7 papers covered on sentiment analysis and earth/planetary system respectively. The remaining of the papers related to vehicle/road, energy, financial, imaging, social media, modulation recognition, radar performance, environmental health, window, machinery, and veterinary.

Although result identified in Table 7 may provide a valuable insight to researchers on the popular categories of SML algorithms used in research previously, but the findings on the

number of the papers by categories may be less accurate in this current study due to the limitation issues especially on the number of databases accessible and short duration selected by the researchers.

Table 7 The classification of usage of SML algorithms based on categories.

No	The classification of the usage of SML algorithms based on categories	Frequency	No	The classification of the usage of SML algorithms based on categories	Frequency
1	Spam/text classification	10	9	Imaging	2
2	Healthcare/medical	9	10	Social media	2
3	Network/software/computer	8	11	Modulation recognition	1
4	Sentiment analysis	7	12	Radar performance	1
5	Earth/planetary system	7	13	Environmental health	1
6	Vehicle/road	4	14	Window	1
7	Energy	3	15	Machinery	1
8	Financial	3	16	Veterinary	1

3.3 RQ2: Comparing the best performance measures of the SML algorithms

With regard to the second research question (RQ2), a comparison between SML algorithms was further conducted in identifying the most performed/best SML model for all 61 articles found in this study.

Prior to that, Figure 3 shows the frequency of SML algorithms used in the articles from highest to lowest count. It can be concluded that the most frequent algorithm used by various researchers in various studies was Support Vector Machine (SVM) with n = 56 (91.80%). SVM is famously used as SML algorithm because this algorithm can be used in the classification and often predicting values with high accuracy. Classification is one of the Data Mining methods which are widely applied in various fields. The algorithm will learn a predictive model based on the relationship between the input variables and the output variable with the availability of a labeled dataset.

This is followed by Naive Bayes (NB) with n = 39, which is 63.93% of the study. There were 32 (52.46%) and 30 (49.18%) articles use Knearest Neighbors (KNN) algorithm and Decision Trees algorithm (DT) respectively in finding the best SML algorithms. Next, Logistic Regression (LR) with n = 24 (39.44%), Random Forest (RF) with n = 23 (37.70%) and Artificial Neural Network (ANN) with n = 20 (32.79%). Other than that, Gradient Boosting Decision Trees/Machine (n = 13, 21.31%) and Other SML algorithm (n = 9, 14.75%). From this finding, the lowest algorithm used by researchers was Linear Regression and Discriminant Analysis (DA) with n = 4 (6.56%) respectively.

Different researchers apply different ML algorithms for their study. Depending on the sample size, attributes available in the dataset, and type of algorithms, the result of the most performed/best SML algorithms are varying from one to another since there are advantages and disadvantages of each algorithms. According to Gorade et al. [73] there are five characteristics which differentiate the quality of the algorithms such as accuracy, computational time to construct the model, strength, data size, and extendibility of the algorithm.

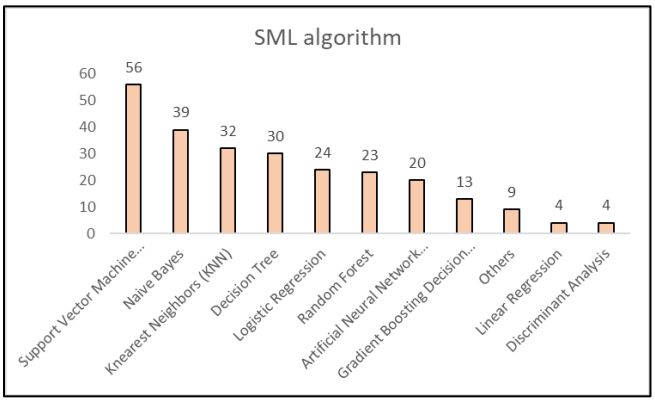

Figure 3 Frequency of SML algorithms.

Table 8 The ranking of the most performed/best SML algorithms.

No	Rank of the most performed/best SML algorithms (Performance)	Frequency	Rank
1	Support Vector Machine (SVM)	18	1
2	Artificial Neural Network (ANN)	10	2
3	Naive Bayes (NB)	8	3
4	Decision Tree (DT)	8	3
5	Random Forest (RF)	8	3
6	Logistic Regression (LR)	5	6
7	Gradient Boosting Decision Trees/Machine	4	7
8	Others	4	7
9	K-nearest Neighbors (KNN)	3	9
10	Linear Regression	2	10
11	Discriminant Analysis (DA)	0	11

Table 8 above presented the frequency and ranking of the most performed/best SML algorithms based on the articles. There are only top 10 algorithms that were taken into the findings such as SVM, ANN, NB, DT, RF, LR, Gradient Boosting Decision Trees/Machine, KNN, Linear Regression, and DA. Other than the stated algorithms were classified into the "Others" category.

In the top 2 ranks, it can be found that 18 articles and 10 articles mentioned SVM and ANN respectively as the best SML algorithm in the prediction studies. This is due to both algorithms successfully classified and predicted values accurately with higher accuracy percentage in the articles. It is also found to be the most appropriate classifier compared to the other algorithms in the stated study. Following that, NB, DT and RF shared the third-ranking place due to 8 articles (differently) stated that these algorithms are the best predictor in the studies. DA is ranked at the last place since none out of 61 articles mentioned that DA is performing as the best SML algorithm.

In conclusion, the results in different performed/best SML algorithms found were unique, highly dependent and influenced by various factors/attributes available in that particular research conducted by researchers.

4.0 LIMITATIONS

Since this study aim was to identify and classify the usage of SML algorithms based on categories and determine and compare the best performance measures of the SML algorithms, hence the research was essentially limited and covering only three digital libraries/databases which were accessed by author. Furthermore, the SLR only include articles that were published in the English Language. There may be a chance that articles in other Language which are relevant to this study was missed.

In the future work, it is recommended to extend this study to access more digital libraries/databases and also to include other than English Language for further broad and in-depth findings. Other than that, the expansions of the study may identify more assessment measures of SML algorithms and include UML.

5.0 CONCLUSIONS

From this SLR study, initially, 305 papers were gathered from online databases from three digital libraries/databases. Based on the exclusion and inclusion criteria in the first phase, there are 137 articles were excluded and after conducting a further assessment on full papers, 76 studies were also excluded and thus left with only 61 papers included in this study. It is interesting to found that SML had been mostly used in classifying spam and text and also in healthcare and medical-related classification research. SVM and ANN are the top two of the best performance measures of the SML algorithms from this study.

REFERENCES

[1] Tzanis, G., Katakis, I., Partalas, I., & Vlahavas, I. (2006). Modern applications of machine learning. *Proceedings of the 1st Annual SEERC Doctoral Student Conference*, 1-10.

[2] Bijamov, A., Shubitidze, F., Fernández, J. P., Shamatava, I., Barrowes, B., & O'Neill, K. (2011). Comparison of supervised and unsupervised machine learning techniques for UXO classification using EMI data. *Proceedings of SPIE - The International Society for Optical Engineering*, 801706, 1-11.

[3] Randle, O. A., Ogunduyile, O. O., Zuva, T., & Fashola, N. A. (2013). A comparism of the performance of supervised and unsupervised machine learning techniques in evolving Awale/Mancala/Ayo game player. *International Journal of Game Theory and Technology*, 1(1), 9-18.

[4] Sathya, R., & Abraham, A. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence*, 2(2), 34-38.

[5] Mallet, R., Hagen-Zanker, J., Slater, R., & Duvendack, M. (2012). The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness*, 4(3), 445-455.

[6] Dallora, A. L., Eivazzadeh, S., Mendes, E., Berglund, J., & Anderberg, P. (2016). Prognosis of dementia employing machine learning and microsimulation techniques: A systematic literature review. *Procedia Computer Science*, 100, 480-488.

[7] Malhotra, R. (2015). A systematic review of machine learning techniques for software fault prediction. *Applied Soft Computing*, 27, 504-518.

[8] Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. d. P., Basto, J. P., & Alcala, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, 1-10.

[9] Azeem, M. I., Palomba, F, Shi, L., & Wang, Q. (2019). Machine learning techniques for code smell detection: A systematic literature review and meta-analysis. *Information and Software Technology*, 108, 115-138.

[10] Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. *Technical Report EBSE 2007-001 Keele University and University of Durham*, 1-57.

[11] Jnr, A. B., & Pa, C. N. (2015). A review on tools of risk mitigation for information technology management. *Journal of Theoretical and Applied Information Technology*, 81, 92-101.

[12] Untawale, T. M., & Choudhari, G. (2019). Implementation of sentiment classification of movie reviews by supervised machine learning approaches. *2019 3rd International Conference on Computing Methodologies and Communication*, 1197-1200.

[13] Moh, M., Gajjala, A., Gangireddy, S. C. R., & Moh, T-S. (2016). On multi-tier sentiment analysis using supervised machine learning. *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 341-344.

[14] Mishu, S. Z., & Rafiuddin, S. M. (2017). Performance analysis of supervised machine learning algorithms for text classification. *2016 19th International Conference on Computer and Information Technology*, 409-413.

[15] Batayev, N. (2019). Gas turbine fault classification based on machine learning supervised techniques. *2018 14th International Conference on Electronics Computer and Computation*, 206-212.

[16] Taher, K. A., Jisan, B. M. Y., & Rahman, M. M. (2019). Network intrusion detection using supervised machine learning technique with feature selection. *2019 International Conference on Robotics, Electrical and Signal Processing Techniques*, 643-646.

[17] Sehgal, A., & Mathur, S. (2019). Plant disease classification using soft computing supervised machine learning. *2019 3rd International Conference on Electronics, Communication and Aerospace Technology*, 75-80.

[18] Dhankhad, S., Mohammed, E., & Far, B. (2018). Supervised machine learning algorithms for credit card fraudulent transaction detection: A comparative study. *2018 IEEE International Conference on Information Reuse and Integration*, 122-125.

[19] Kaur, V. D. (2019). Sentimental analysis of book reviews using unsupervised semantic orientation and supervised machine learning approaches. *2018 2nd International Conference on Green Computing and Internet of Things*, 519-524.

[20] Navaney, P., Dubey, G., & Rana, A. (2018). SMS spam filtering using supervised machine learning algorithms. *2018 8th International Conference on Cloud Computing, Data Science and Engineering*, 43-48.

[21] Bari, M., Taher, H., Sherazi, S. S., & Doroslovacki, M. (2017). Supervised machine learning for signals having RRC shaped pulses. *2016 50th Asilomar Conference on Signals, Systems and Computers*, 652-656.

[22] Vyas, T., Prajapati, P., & Gadhwal, S. (2015). A survey and evaluation of supervised machine learning techniques for spam e-mail filtering. *2015 IEEE International Conference on Electrical, Computer and Communication Technologies*, 1-7.

[23] Bhamare, D., Salman, T., Samaka, M., Erbad, A., & Jain, R. (2017). Feasibility of supervised machine learning for cloud security. *2016 International Conference on Information Science and Security*, 1-5.

[24] Aziz, A. A., Starkey, A., & Bannerman, M. C. (2018). Evaluating cross domain sentiment analysis using supervised machine learning techniques. *2017 Intelligent Systems Conference*, 689-696.

[25] Soheily-Khah, S., Marteau, P-F., & Bechet, N. (2018). Intrusion detection in network systems through hybrid supervised and unsupervised machine learning process: A case study on the ISCX dataset. *2018 1st International Conference on Data Intelligence and Security*, 219-226.

[26] Habib, A., Akbar, S., Asghar, M. Z., Khattak, A. M., Ali, R., & Batool, U. (2018). Rumor detection in business reviews using supervised machine learning. *2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing*, 233-237.

[27] Alkalbani, A. M., Ghamry, A. M., Hussain, F. K., & Hussain, O. K. (2016). Predicting the sentiment of SaaS online reviews using supervised machine learning techniques. *2016 International Joint Conference on Neural Networks*, 1547-1553.

[28] Li, W., & Meng, W. (2015). An empirical study on email classification using supervised machine learning in real environments. *2015 IEEE International Conference on Communications*, 7438-7443.

[29] Tudu, R., Saha, S., Pritam, P. N., & Palit, R. (2019). Performance analysis of supervised machine learning approaches for Bengali text categorization. *2018 5th Asia-Pacific World Congress on Computer Science and Engineering*, 221-226.

[30] Pan, Y., Zhang, J., Luo, G. Q., & Yuan, B. (2018). Evaluating radar performance under complex electromagnetic environment using supervised machine learning methods: A case study. *2018 8th International Conference on Electronics Information and Emergency Communication*, 206-210.

[31] Khadse, V., Mahalle, P. N., & Biraris, S. V. (2019). An empirical comparison of supervised machine learning algorithms for Internet of Things data. *2018 4th International Conference on Computing Communication Control and Automation*, 1-6.

[32] Smadi, M., & Qawasmeh, O. (2018). A supervised machine learning approach for events extraction out of Arabic tweets. *2018 5th International Conference on Social Networks Analysis, Management and Security*, 114-119.

[33] Makary, M., Oakes, M., Mitkov, R., & Yammout, F. (2017). Using supervised machine learning to automatically build relevance judgments for a test collection. *2017 28th International Workshop on Database and Expert Systems Applications*, 108-112.

[34] Abd, D. H., & Al-Mejibli, I. S. (2019). Monitoring system for sickle cell disease patients by using supervised machine learning. *2017 2nd Al-Sadiq International Conference on Multidisciplinary in IT and Communication Science and Applications*, 119-124.

[35] Motamedidehkordi, N., Amini, S., Hoffmann, S., Busch, F., & Fitriyanti, M. R. (2017). Modeling tactical lane-change behavior for automated vehicles: A supervised machine learning approach. *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems*, 268-273.

[36] Herath, H. M. M. G. T., Kumara, J. R. S. S., Fernando, M. A. R. M., Bandara, K. M. K. S., & Serina, I. (2018). Comparison of supervised machine learning techniques for PD classification in generator insulation. *2017 IEEE International Conference on Industrial and Information Systems*, 1-6.

[37] Panahandeh, G., Ek, E., & Mohammadiha, N. (2017). Road friction estimation for connected vehicles using supervised machine learning. *2017 IEEE Intelligent Vehicles Symposium (IV)*, 1262-1267.

[38] Zerrouki, N., Harrou, F., Houacine, A., & Sun, Y. (2017). Fall detection using supervised machine learning algorithms: A comparative study. *2016 8th International Conference on Modelling, Identification and Control*, 665-670.

[39] Sibaroni, Y., Widyantoro, D. H., & Khodra, M. L. (2017). Extend relation identification in scientific papers based on supervised machine learning. *2016 International Conference on Advanced Computer Science and Information Systems*, 379-384.

[40] Samal, B. R., Behera, A. K., & Panda, M. (2017). Performance analysis of supervised machine learning techniques for sentiment analysis. *2017 3rd International Conference on Sensing, Signal Processing and Security*, 128-133.

[41] Zhou, DX. (2015). Machine learning algorithms. *Encyclopedia of Applied and Computational Mathematics*, 839-841.

[42] Win, K. Y., Maneerat, N., Choomchuay, S., Sreng, S., & Hamamoto, K. (2019). Suitable supervised machine learning techniques for malignant mesothelioma diagnosis. *2018 11th Biomedical Engineering International Conference*, 1-5.

[43] Aboueata, N., Alrasbi, S., Erbad, A., Kassler, A., & Bhamare, D. (2019). Supervised machine learning techniques for efficient network intrusion detection. *2019 28th International Conference on Computer Communication and Networks*, 1-8.

[44] Elkanishy, A., Badawy, A. A., & Furth, P. M. (2019). Supervising communication SoC for secure operation using machine learning, *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems*, 582-585.

[45] Putra, I. P. E. S., & Vesilo, R. (2018). Window-size impact on detection rate of wearable-sensor-based fall detection using supervised machine learning. *2017 IEEE Life Sciences Conference*, 21-26.

[46] Lushan, M., Bhattacharjee, M., Ahmed, T., Rahman, M. A., & Ahmed, S. (2019). Supervising vehicle using pattern recognition: Detecting unusual behavior using machine learning algorithms. *2018 IEEE Region Ten Symposium*, 277-281.

[47] Gupta, S., Mittal, M., & Padha, A. (2018). Predictive analytics of sensor data based on supervised machine learning algorithms. *2017 International Conference on Next Generation Computing and Information Systems*, 171-176.

[48] Vergara-Reyes, J., Martinez-Ordonez, M. C., Ordonez, A., & Rendon, O. M. C. (2017). IP traffic classification in NFV: A benchmarking of supervised machine learning algorithms. *2017 IEEE Colombian Conference on Communications and Computing*, 1-6.

[49] Gupta, M., & Gupta, B. (2018). A comparative study of breast cancer diagnosis using supervised machine learning techniques. *2018 2nd International Conference on Computing Methodologies and Communication*, 997-1002.

[50] Robles-Durazno, A., Moradpoor, N., McWhinnie, J., & Russell, G. (2018). A supervised energy monitoring-based machine learning approach for anomaly detection in a clean water supply system. *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 1-8.

[51] Zhang, Y., & Ma, Y. (2019). Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia. *Computers in Biology and Medicine*, 106, 33-39.

[52] Bao, W., Lianju, N., & Yue, K. (2019). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Systems with Applications*, 128, 301-315.

[53] Mukhtar, N., Khan, M. A., & Chiragh, N. (2018). Lexicon-based approach outperforms supervised machine learning approach for urdu sentiment analysis in multiple domains. *Telematics and Informatics*, 35(8), 2173-2183.

[54] Higaki, S., Miura, R., Suda, T., Andersson, L. M., Okada, H., Zhang, Y., Itoh, T., Miwakeichi, F., & Yoshioka, K. (2019). Estrous detection by continuous measurements of vaginal temperature and conductivity with supervised machine learning in cattle. *Theriogenology*, 123, 90-99.

[55] Van Zoonen, W., & Van Der Meer, T. G. L. A. (2016). Social media research: The application of supervised machine learning in organizational communication research. *Computers in Human Behavior*, 63, 132-141.

[56] Ozbas, E. E., Aksu, D., Ongen, A., Aydin, M. A., & Ozcan, H. K. (2019). Hydrogen production via biomass gasification, and modeling by supervised machine learning algorithms. *International Journal of Hydrogen Energy*, 44(32), 17260-17268.

[57] Rahman, I., Kuzlu, M., & Rahman, S. (2018). Power disaggregation of combined HVAC loads using supervised machine learning algorithms. *Energy and Buildings*, 172, 57-66.

[58] Lynch, C. M., Abdollahi, B., Fuqua, J. D., de Carlo, A. R., Bartholomai, J. A., Balgemann, R. N., van Berkel, V. H., & Frieboes, H. B. (2017). Prediction of lung cancer patient survival via supervised machine learning classification techniques. *International Journal of Medical Informatics*, 108, 1-8.

[59] McAllister, P., Zheng, H., Bond, R., & Moorhead, A. (2018). Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. *Computers in Biology and Medicine*, 95, 217-233.

[60] Bostik, O., & Klecka, J. (2018). Recognition of CAPTCHA characters by supervised machine learning algorithms. *IFAC PapersOnLine*, 51(6), 208-213.

[61] Oonk, S., & Spijker, J. (2015). A supervised machine-learning approach towards geochemical predictive modelling in archaeology. *Journal of Archaeological Science*, 59, 80-88.

[62] Belavagi, M. C., & Muniyal, B. (2016). Performance evaluation of supervised machine learning algorithms for intrusion detection. *Procedia Computer Science*, 89, 117-123.

[63] Lanka, P., Rangaprakash, D., Dretsch, M. N., Katz, J. S., Denney, T. S., & Deshpande, G. (2019). Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets. *Brain Imaging and Behavior*, 1-39.

[64] Di Noia, A., Martino, A., Montanari, P., & Rizzi, A. (2019). Supervised machine learning techniques and genetic optimization for occupational diseases risk prediction. *Soft Computing*, 24, 4393-4406.

[65] Verma, H., & Verma, G. (2019). Prediction model for bollywood movie success: A comparative analysis of performance of supervised machine learning algorithms. *The Review of Socionetwork Strategies*, 14, 1-17.

[66] Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review*, 52(1), 273-292.

[67] Mwanga, E. P., Mapua, S. A., Siria, D. J., Ngowo, H. S., Nangacha, F., Mgando, J., Baldini, F., Jimenez, M. G., Ferguson, H. M., Wynne, K., Selvaraj, P., Babayan, S. A., & Okumu, F. O. (2019). Using mid-infrared spectroscopy and supervised machine-learning to identify vertebrate blood meals in the malaria vector, Anopheles arabiensis. *Malaria Journal*, 18(1), 1-9.

[68] Rovini, E., Maremmani, C., Moschetti, A., Esposito, D., & Cavallo, F. (2018). Comparative motor pre-clinical assessment in parkinson's disease using supervised machine learning approaches. *Annals of Biomedical Engineering*, 46(12), 2057-2068.

[69] Miettinen, O. (2018). Protostellar classification using supervised machine learning algorithms. *Astrophysics and Space Science*, 363(9), 197, 1-15.

[70] Arango, R. B., Díaz, I., Campos, A., Canas, E. R., & Combarro, E. F. (2016). Automatic arable land detection with supervised machine learning. *Earth Science Informatics*, 9(4), 535-545.

[71] Aljaaf, A. J., Al-Jumeily, D., Hussain, A. J., Fergus, P., Al-Jumaily, M., & Radi, N. (2015). A systematic comparison and evaluation of supervised machine learning classifiers using headache dataset. *International Conference on Intelligent Computing: Advanced Intelligent Computing Theories and Applications*, 101-108.

[72] Prusty, M. R., Chakraborty, J., Jayanthi, T., & Velusamy, K. (2015). Performance comparison of supervised machine learning algorithms for multiclass transient classification in a nuclear power plant. *5th International Conference on Swarm, Evolutionary, and Memetic Computing*, 111-122.

[73] Gorade, S., Deo, A., & Purohit, P. (2017). A study of some data mining classification techniques. *International Research Journal of Engineering and Technology*, 4(4), 3112-3115.