Supervised Machine Learning a Brief Survey of Approaches

Esraa Najjar[*], Aqeel Majeed Breesam[**]

[*] Computer Science Department, General Directorate of Education in Najaf Governorate, Al-Najaf, Iraq
Email: esraas.najjar@uokufa.edu.iq
https://orcid.org/0009-0005-1800-3618

[**] Institute of Medical Technology / Baghdad, Middle Technical University, Baghdad, Iraq
Email: aqeelmajeed@mtu.edu.iq
https://orcid.org/0000-0003-3735-9028

Abstract

Machine learning has become popular across several disciplines right now. It enables machines to automatically learn from data and make predictions without the need for explicit programming or human intervention. Supervised machine learning is a popular approach to creating artificial intelligence. A computer algorithm is trained on input data that has been labeled for a certain output, making it one of two major areas of machine learning that has seen a lot of successful research. The model is trained until it can identify the underlying correlations and patterns between the input and output labels, enabling it to generate accurate labeling results when confronted with unexplored data. Supervised learning is good at solving classification and regression problems. The problem of regression occurs when the outputs are continuous, while the problem of classification occurs when the outputs are categorical. We will concentrate on the benefits and drawbacks of supervised learning algorithms in this review. Creating a precise model of the distribution of class labels in terms of predictor features is the aim of supervised learning. This work studied the most popular supervised machine learning methods, including Naive Bayes, Decision Trees, Support Vector Machines, Logistic Regression, K-Nearest Neighbors, and Deep Learning, which were discussed in this paper. We also emphasized the algorithms' advantages and disadvantages, and we ultimately talked about the difficulties in developing supervised machine learning algorithms.

Keywords- Machine learning, Supervised learning, Accuracy, Classification Algorithms, Classifiers.

I. INTRODUCTION

Machine learning is a branch of computer science that looks for patterns in data to improve performance on a range of activities [1]. Information technology, statistics, probability theory, artificial intelligence, psychology, neurology, and many more fields are all heavily reliant on machine learning. Building a model that depicts a well chosen data set using machine learning is all that is required to solve the issue [2]. As a result of the growth of machine learning and its blossoming in imitation of the human brain, statistics has evolved into a broad subject that provides the fundamental statistical computational theories for learning processes [3].

Machine learning has been used in several fields, including robotics, natural language processing, computer games, data mining, virtual personal assistants (like Google), pattern recognition, traffic forecasting, online traffic networks, product recommendations, stock market predictions, medical diagnostics, online fraud predictions, farming advice, Email spam filtering and a social media service (face recognition on Facebook) add to crime prediction using video surveillance systems[4].
Machine learning approaches primarily fall into two categories [5] as Figure 1.

- Supervised Learning
- Unsupervised Learning

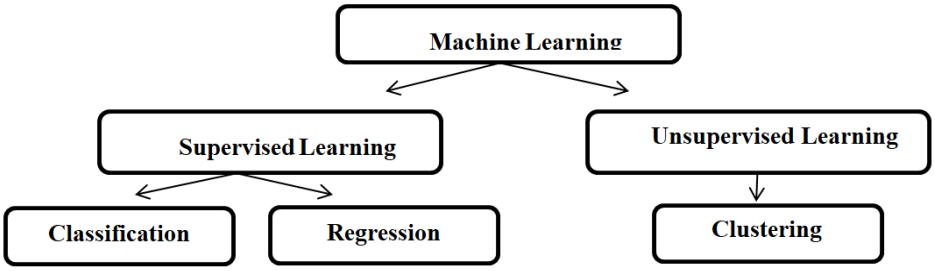

Fig. 1. Types of Machine Learning [5]

II. SUPERVISED LEARNING

It is based on a labeled dataset, the model receives labels as it is being developed. These labeled datasets undergo training in order to produce significant results during the decision-making process. The choice and extraction of the particular feature set to be employed are crucial to the effectiveness of both learning techniques. Two types of data are needed to use machine learning techniques: a training set and a test set [6].

Computers cannot learn from prior experience; this ability belongs to humans only. Our primary goal in supervised or inductive machine learning is to discover an objective job that is used to forecast class values [7]. As demonstrated below in Figure 2, this is how supervised machine learning is applied to real-world issues [8].

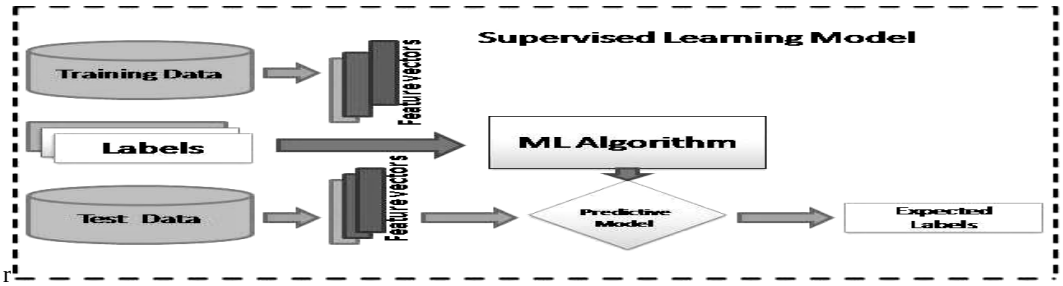

Fig. 2. Supervised Machine Learning Model [9]

There are a few steps that must be taken during the learning process in supervised machine learning. Data collection is the first step in this process, and it is the process of gathering all necessary and relevant elements that will be employed in this process. Preprocessing the data is the second stage, and several methods have been employed to manage and recreate deleted or missing data [10][11]. Finding and removing as many unused and superfluous features as you can is the process of feature subset selection. This lowers the dimensionality of the data and makes it possible for data mining techniques to work more quickly and efficiently [12]. The most common way of choosing the best algorithm is an imperative and huge step that is resolved once testing has ended up being satisfactory and adequate and a classifier is accessible for utilization. Many studies have provided a comparison of supervised learning algorithms as table1.

Here are common real-world examples for Supervised Machine Learning:

- Stock price prediction
- Image identification
- Customer churn prediction
- Spam detection
- Weather forecasting

Table 1. Example for Supervised Machine Learning

Research	Period	Title	Dataset	Methodology	Main conclusion	Strength	Weakness
"P. Buczak, H. Huang, B. Forthmann, and P. Doebler"[14]	2023	"The machines take over: A comparison of various supervised learning approaches for automated scoring of divergent thinking tasks"	existing data from two distinct laboratories	Random Forest and Boost tended, Support Vector Regression	In comparison to Random Forest and Boost, Support Vector Regression consistently outperformed the other two cutting-edge machine learning algorithms.	Constructed a number of prediction models utilizing features that reflected the response's meta data or features obtained from word embeddings that were collected using previously trained GloVe and Word2Vec word vector spaces.	It demonstrates that the models did not adapt successfully to a new sample.
"E. Dritsas and M. Trigka" [17]	2023	"Supervised Machine Learning Models for Liver Disease Risk Prediction"	Based on the Indian Liver Patients' Records dataset. There are 579 individuals in the particular dataset, 439 of whom are men (75.2%) and 140 of whom are women (24.2%).	There are many machine learning models, including Stacking, Bagging, Voting, RF, RT, J48, RotF, RepTree, AdaBoostM1, SVM, ANN, kNN, NB, and LR.	The key thesis of this study is that the Voting classification approach surpasses the others with an SMOTE with 10-fold cross-validation, with an Accuracy, Recall, and F-measure of 80.1%, a Precision of 80.4%, and an AUC of 88.4%.	Comparing our suggested model's (i.e., Voting) accuracy to similar previous research studies using the same attributes as our dataset, we find that our model is more accurate.	Utilize solely machine learning
"Poonam Chaudhari, Himanshu Agarwal, Vikrant Bhateja"[15]	2019	"Data augmentation for cancer classification in oncogenomics: an improved KNN based approach"	The NCBI repository databases that contained the gene expression microarray data utilized in this study were accessible to the general public.	Using classifiers like SVM, J48 KNN	When using raw data as input to the classifiers, The classifiers' average classification accuracy outperforms the traditional KNN approach by 16% and 7.72%, respectively.	The suggested approach accomplishes two goals: first, it ensures data sensitivity for important applications and, second, it improves classification accuracy.	Gene expression data are dominated by a limited number of samples and a large number of characteristics.
"J. Xu, H. Mu, Y. Wang, and F. Huang"[16]	2018	**"**Feature Genes Selection Using Supervised Locally Linear Embedding and Correlation Coefficient for Microarray Classification"	The simulation experiment uses four publicly available tumor microarray datasets	Using classifiers like SVM ,NB, KNN	Genes that are redundant are removed using supervised LLE. Spearman's rank correlation coefficient eliminates the coexpression relationship between genes when taking the qualities into account.	The concept of locally linear embedding and correlation coefficient algorithms is the foundation of the effective feature selection technique known as supervised locally linear embedding and Spearman's rank correlation coefficient (SLLE-SC2)..	The majority of current approaches perform poorly in terms of classification and have a high time complexity.

Source: Our own evaluation

III. ALGORITHM SELECTION

The choice of algorithm is one of the important steps to achieving good results. The score of the classifier is mainly built on accuracy and ability, which are measured by the following formula Eq (1)[18]:

$$\text{Accuracy} = \frac{\text{Number of correct classification}}{\text{Total of number test cases}} \quad (1)$$

Researchers use a number of different techniques to determine a classifier's accuracy. The practice set is divided by some researchers, employing one-third for performance assessment and two-thirds for training. Another strategy is rotation estimation or cross-

validation (CV). CV offers a technique to make greater use of the examples that are already available. We divide the training sample into k disjoint groups of equal size in a k-fold cross-validation procedure [19].

A model is then created from each model using a learning algorithm, and its effectiveness is assessed. The ultimate performance is calculated using the average performance of all these models. Keep in mind that when k equals the quantity of objects in the training sample, the strategy is known as omission. For computational reasons, small values of k (10 or 20) are typically recommended [20]. The accuracy of trained classifiers on certain datasets can be statistically compared using comparisons of supervised ML techniques. It can compare the accuracy of every pair of classifiers on a significant test set by applying two different algorithms to samples of training sets of size N. [21]. In the following, we will be discussing specifically the most important supervised machine learning algorithms.

1. Naive Bayes (NB)

The Bayes' theorem-based statistical algorithm NB can be successfully trained in a supervised learning environment. NB is effective enough to classify in many domains[22] and has been used to classify documents over the decades. The Naïve Bayesian algorithm gives better results in solving complex problems[23]. When NB employs the Bayes'rule to determine the probable class c* for a new document d, it calculates in Eq (2):

$$c* = \text{argmax}_c \, P(c \mid d) \tag{2}$$

The Bayes rule is applied by the NB classifier in Eq (3):

$$P(c \mid d) = (P(c)\, P(d \mid c)) \,/\, P(d) \tag{3}$$

The primary theoretical problem of NB it is that by assuming the conditional independence of features fi's given d's class, P(d) plays no impact in selecting c* for the estimate of the term P(dc) as Eq (4):

$$P_{NB(c\backslash d)} = \frac{p(c)(\prod_{i=1}^{m} p(fi\backslash c)^{ni(d)}}{P(d)} \tag{4}$$

Using add-one smoothing, the training approach calculates the relative frequency of P(c) and P(fi | c)[24]. NB classifier Although the conditional independence assumption is false in reality, it nonetheless works effectively for sentiment classification[25][26].

2. Decision Trees

A decision tree responds to the machine learning problem by representing the data as a tree. The decision tree is a graph that displays options and their results in a structure like a tree. The edges of the graph's nodes serve as the requirements or rules for making decisions, while the nodes themselves represent an event or a choice. In every tree, there are nodes and branches. Each node represents a set of traits that need to be categorized, and each branch indicates a potential value for the node [27].
Different types can be used the algorithms are:[28].

- ID3(Iterative Dichotomiser3)
- C4.5 (an ID3 successor).
- CART, short for Classification and Regression Trees

Figure 3 shows a straightforward model of a decision tree having a single binary target variable, Y (0 or 1), two continuous variables, x1 and x2, and a range of 0 to 1. A decision tree model is primarily made up of nodes and branches, whereas the main modeling actions are splitting, stopping, and pruning[29].

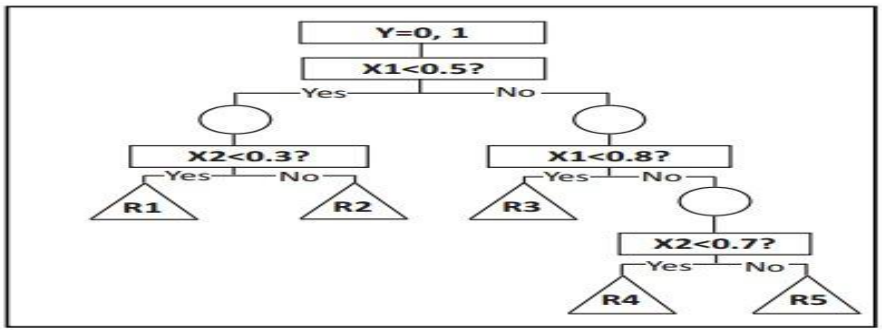

Fig. 3. Example of a binary target variable-based decision tree [29]

Three types of nodes are possible. (a) A root node, also known as a decision node, represents the choice to divide all data into two or more sets that are incompatible. (b) Internal nodes, also referred to as chance nodes, represent one of the potential outcomes that might be possible at that exact point in the tree structure; the node's top edge is connected to its parent node, and its bottom edge is connected to its child nodes or leaf nodes. (c) Leaf nodes, also known as end nodes, are the outcome of a set of choices or occurrences. Branch nodes represent occurrences or random events that originate from internal and root nodes. To establish a hierarchy of branches[29]. These 'if-then' rules can also be used to the expression of these decision tree routes. For example, "outcome j occurs if condition(s) 1

and 2 and... and condition(s) k) exist. Only input variables pertinent to the target variable are used to split parent nodes into purer child nodes of the target variable[30].

It is successful in both classification and regression applications because it can be used to forecast both continuous and discrete data. Decision trees are easy to understand because of how straightforward they are as an algorithm to imagine and infer. Since it can identify nonlinear correlations, it can be used to categorize data that cannot be separated linearly. When working with non-linear data, decision trees have the benefit of not requiring feature transformation because they do not consider numerous weighted combinations simultaneously [31].

3. Support Vector Machines (SVM)

The basis of SVM was invented by Vladimir Vapnik and his group at AT&T Bell Laboratories [32]. Among the well-known data mining approaches, it is regarded as one of the most well-liked algorithms in supervised machine learning and one of the most potent, reliable, and accurate procedures[33].

A simple explanation is that whenever a collection of training examples is provided, each example is categorized into one of two groups. A model is developed by the SVM training algorithm to determine whether a new sample belongs to one class or another. In an intuitive way. SVMs map sample points in space so that instances from various classes are as far apart as feasible. This is how they represent samples. Based on whichever side of the space a fresh sample falls on, they are then mapped to the same side and projected to belong to a class [5].

More formally, SVMs are classified using a hyperplane or set of hyperplanes that create separation between data points with the aid of support vectors. Figure 4 shows the concept of a maximum margin hyperplane separating positive examples (green squares) from negative examples (red circles); the darker green squares and red circles represent associated support vectors. A complete separation is attained by the hyperplane with the greatest functional margin from any class's closest training data point [34] [35].

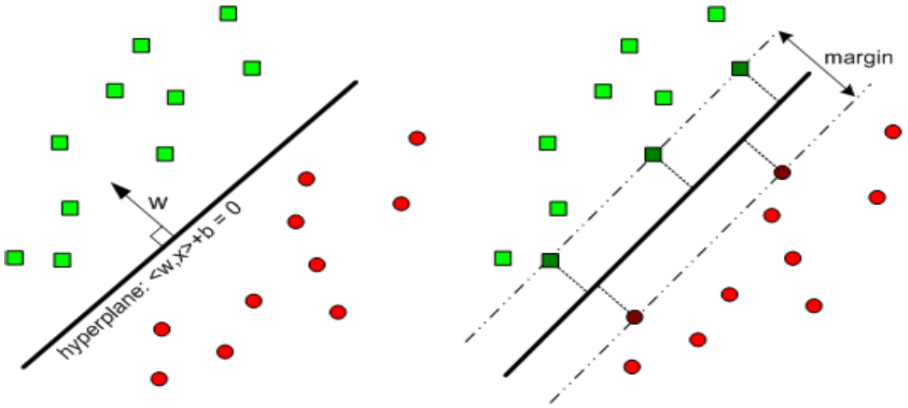

Fig. 4. Maximum margin [36]

Some researchers' view of SVM is that they get an optimal hyperplane as the solution to the learning problem. For the output of linear SVMs, a separating hyperplane can be expressed as a down Eq.(5) If x is a feature vector of classification text made up of words, w is the weight of identical x, b is the bias parameter [37], as shown in Figure 5.

$$z = w.x + b \quad (5)$$

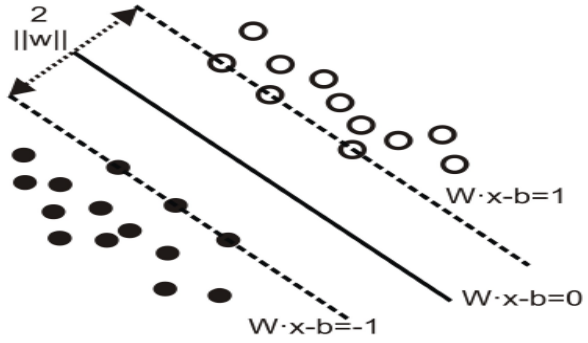

Fig. 5. A hyperplane and the support vectors [38]

4. Logistic Regression

The regression fits the data to a logistic function to forecast the likelihood of an event. A set of predictor variables that could otherwise be numerical or categorical are used in logistic regression, just like any other type of regression analysis[3]. This algorithm is employed in many different domains, such as supervised machine learning, most medical specialties, and social sciences[39].

Similar to NB, logistic regression works by taking some weighted input points, calculating the logarithms, and adding them linearly, weighting each function to increase its capacity, and then adding the results. Logistic regression and Bayesian naive are distinguished by the fact that the latter is a generative classifier while the former is a recognizable classifier that can predict outcomes using logistic regression modeling when the outcome variable is binary. The definition of the logistic regression hypothesis in Eq (6)

$$h\_\theta = g(\theta\text{^}tx) = \text{sigmoid}(Z) \quad (6)$$

Where: sigmoid (z)= 1 / (1+e^(-z)).

Figure 6 illustrates the characteristic features of the sigmoid function that give values in the [0, 1] range.

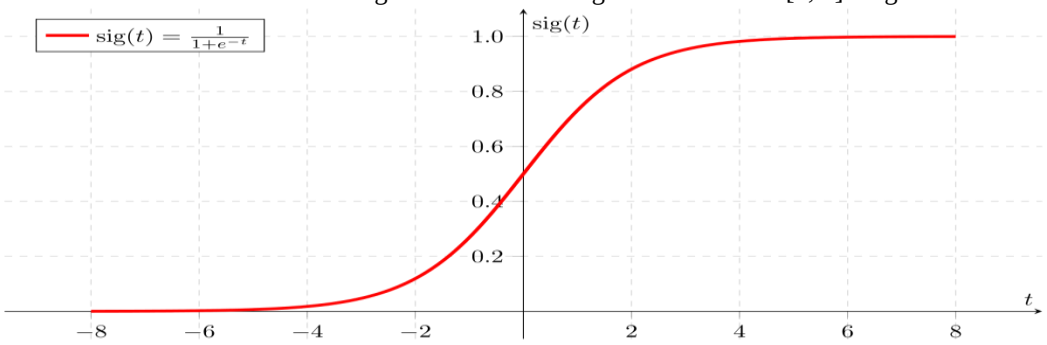

Fig. 6. Visual representation of the Logistic Function Y(predicted) becomes 1 when Z reaches infinity, while Y(predicted) becomes 0 when Z reaches negative infinity [40]

When the dataset can be separated linearly, logistic regression performs well. In high-dimensional datasets, logistic regression is less likely to overfit, although it still has the potential to do so. In addition to measuring the amount of the predictors' connection (coefficient magnitude), logistic regression also indicates whether the correlation is positive or negative[4].

5. K-Nearest Neighbors(KNN)

KNN stores all available states and corrects the state based on a similarity measure, for example, distance functions. For that, think about a straightforward approach it can be used to solve problems involving both classification and regression. The basis of its operation is to be used in statistical statistics and to recognize patterns based on their nearest neighbors, which must be an even number. Uses distance factors such as Euclidean, Manhattan, Minkowski, etc.[41].

The KNN() function by default uses the Euclidean distance, which can be determined using the Eq (7) below.

$$D(p,q) = \sqrt{(p1+q2)^2 + (p2+q2)2 + \cdots + (pn-qn)^2} \quad (7)$$

Where p and q are the variables being compared to the n characteristics. There are several methods to calculate distance, such as utilizing the Manhattan distance[42].

Most of the time, we only used K as an odd number when making decisions. Assigned to the class of its closest neighbor if K = 1. If K is an odd number, the class with the most votes from its closest neighbor will be assigned [43].

6. Deep Learning

Deep learning is a branch of supervised machine learning that uses multiple layers of neural networks to handle and perform calculations on large amounts of data[44]. A deep learning algorithm functions by taking into account the potential and operation of the human brain. The deep learning algorithms may be applied to both structured and unstructured types of information since they can learn without human supervision [45]. Deep learning can be used for a variety of businesses, inclusion healthcare, banking, finance, and the Internet [46], similar to how the human brain employs millions of neurons to calculate information, deep learning algorithms rely on neural networks in order to function. There are supervised deep learning algorithms.

6.1 Artificial Neural Network (ANN):

Like the human brain, artificial neural networks function. The human brain has billions of neurons, each of which is made up of a cell body that is responsible for processing input by sending it to neurons that are stowed away and producing the final output. Classification issues, time series, and function approximation are just a few of the hard jobs that an ANN can perform with ease. It is a flexible and strong mathematical instrument[47].
An ANN can perform better by drawing from its prior experiences, which is its highest capacity. An ANN initially recognizes patterns during the training stage using inputs provided to the input layer. Ann compares the output to the actual outcome at this point, and she recognizes the difference between the two as a mistake [48].
The goal is to limit the blunder by changing the weight and inclination of the correlation, which is known as backpropagation. With the course of back engendering, the distinction between the ideal result and the genuine result delivers the least blunder [49].
An artificial neural network is made up of three layers of neurons: an input layer (or node, unit), one to three hidden layers, and a layer of output neurons at the top. A typical architecture is shown in Figure 7 by the lines that link the neurons. There is a weight for each connection, which is a count of numbers. Neuron I's output in the hidden layer as Eq (8).

$$h_i = \sigma \sum_{J=1}^{N} V_{ij}\ X_J\ + T_i^{hid} \tag{8}$$

N stands for the overall number of input neurons, Vij stands for the weights, xj stands for the input neurons, and Tihid stands for the threshold terms for the hidden neurons. Furthermore, rather than paralyzing the brain network with divergent neurons, the activation function seeks to create nonlinearity into the neural network in order to impose a value limit [48].

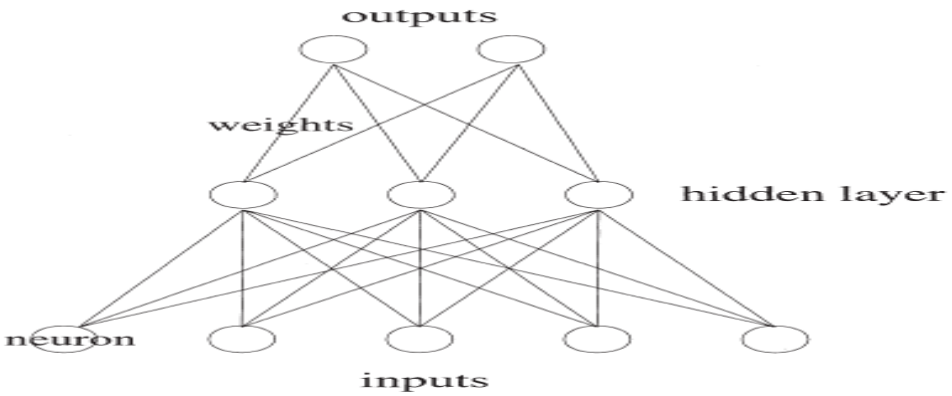

Fig. 7. Architecture of a neural network [48]

6.2 Convolutional Neural Network (CNN):

CNN contains different layers that process and extract signified features from the image [50]. This network's neurons have predispositions and learnable weights. Every neuron point receives the input, processes it with a dot product, and then alternately responds to it with nonlinearity. Picture characterization is the most common way of tolerating an information picture and producing an output class (a feline, a canine, and so on) or a likelihood of classes that best match the picture [51].

CNNs were first broadcast in 1979, but they only became well-known in 2012 after CNNs decisively defeated all other networks in the renowned ImageNet competition. Three convolutional layers and three pooling layers, and five convolutional layers made up the network known as Alex Net. layers that are completely joined, as seen in Figure [52].

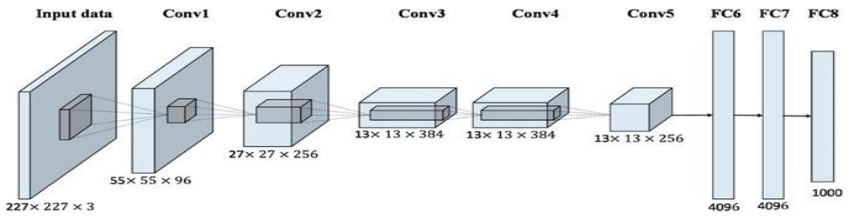

Fig. 8. Convolutional neural network architecture of Alex Net [52]

6.3 Recurrent Neural Network (RNN):

The RNN algorithm is regarded as being most suited for sequential data and is most frequently used in time-series analysis, machine translation, natural language processing, handwriting recognition, and picture captioning. The hidden state, which memorizes certain information about a sequence, is the most crucial component of RNNs[53].

RNNs are now capable of spotting patterns in sequential data, such as audio, video, and text. The long short-term memory (LSTM) network enhances pattern identification as a more contemporary and complicated RNN. Figure 9 displays the RNN and LSTM unit topologies [54].

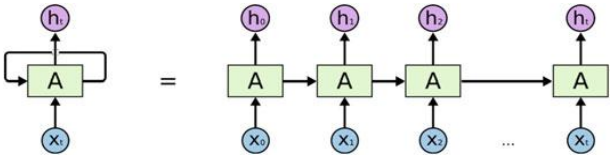

Fig. 9. A single recurrent unit's RNN architecture [52]

IV. IMPORTANCE OF ETHICS FOR SUPERVISED LEARNING

Machine learning (ML) and deep learning (DL) have advanced significantly over the past five years and are currently used in a variety of fields, such as agriculture, animal science, plant science, healthcare, and education. Examining the ethical implications of AI at this point is crucial along with its expansion. Researchers contribute significantly to the governance and ethics of AI through their work, advocacy, openness, and employment decisions.

V. CHALLENGES OF SUPERVISED LEARNING

Many supervised machine learning techniques have high computational complexity and effectiveness, particularly when dealing with huge datasets (e.g., memory limitations; runtimes of a few hours to a few weeks). Label difficulty, not all labels are created equal, just as not all projects are equally difficult depending on the area of application. Incorrect labeling: since people are fallible, mistakes will inevitably be found in data. In evaluations of training and test set quality, it is still very difficult to measure training set quality objectively because the building of training sets is still a fairly ad hoc procedure [55]. We discussed the benefits and drawbacks of algorithms in Table 2.

Table 2. Advantages and Disadvantages of different classifications.

Name of the Classification	Advantages	Disadvantages
Naive Bayes	The speed at which they can be trained and queried with large datasets, relative simplicity, and flexible method for dealing with each number of attributes or classes[28].	The hypothesis of independent features exists, Data scarcity[56] The hypothesis of independent features exists. Practically speaking, it is very unlikely that the model will produce a collection of predictors that are 100 percent independent. Zero posterior probability results when a training tuple for a certain class does not exist. The model is unable to predict this situation. The term "zero probability/frequency problem" refers to this issue[57].
Decision Trees	Selection of features or variable screening. Data preparation is simple[31].	The time complexity appropriate to run this process is huge as the number of records keeps increasing as the decision tree increases with scalar variables it takes a lot of time to train[58]. Refusal to handle continuous data, it very difficult to deal with missing data, they are prone to sampling errors because they are overfitting[28].
Support Vector Machines	It builds very accurate classifiers and effective ones in high dimensional spaces, requires less over-fitting, is widely used in many applications, is memory-efficient in using a flexible decision	A binary classifier: That is, pair-wise classifications (one class against all others, for all classes) can be used to achieve a multi-class classification and it is

	function termed (support vectors) that uses a subset of training points Different kernel tasks may be unique to the decision function. [59].	computationally expensive[60].
Logistic Regression	Simple to comprehend and interpret. It can be used for both binary and multi-class classification problems because it is effective and quick to compute, suitable for large datasets, and provides probability scores for the predicted outcomes, which can be interpreted as the likelihood, that the outcome will occur. It also handles categorical and continuous predictor variables[61].	Nonlinear problems cannot be resolved using logistic regression because it has a linear decision surface, Data that cannot be separated linearly. Complex relationships are difficult for logistic regression to capture. Neural networks and other powerful and complex algorithms can quickly outperform this method[62]. Only significant and pertinent features should be used when developing the model to prevent the probabilistic predictions made by the model from being inaccurate and the anticipated value of the model from decreasing.[41] [61].
K-Nearest Neighbors	Easy to implement, easy to understand. Nonlinear decision limits can be learned for classification and regression. It is possible to have a very flexible decision border that can adjust the value of K.No for training time for classification /regression. It has one hyperparameter: There is only one hyperparameter, the K-value. This makes tuning the hyperparameter easy.[63].	The time complexity and high prediction complexity for huge datasets. Sensitive to outliers: The classification interval can be altered by a single incorrectly tagged case. This could be more of an issue, particularly at larger dimensions[64].
Convolutional Neural Network	Improvements in precision observed in the classification/prediction problems, CNNs can automatically learn relevant features from the input data, CNNs learn a hierarchical representation of the input data[65]	Computational Complexity, CNNs can be difficult to interpret. CNNs can be prone to overfitting, The training process takes a long time if CNN has many levels[65].

Source: Our own evaluation

VI. CONCLUSION

Supervised machine learning methods are used in various fields. The ability to collect data and produce data output based on previous experiences is one of the benefits of guided supervised machine learning under supervision. Experience is used to help optimize performance standards. Many problems with computation in the real world can be solved with the help of assisted supervised machine learning. The selection of algorithms in machine learning (ML) is mostly determined by the nature of the problem. In this study, we provided a variety of methods utilized in supervised algorithms.

We noticed that the most used algorithms are deep learning, which is a potent classifier and has quickly gained popularity in a wide range of application fields over the last few years; SVM, which builds very accurate and effective classifiers in high-dimensional spaces that require less over-fitting; and versatile, while NB may require a relatively minimal dataset, relative simplicity, as well as a versatile method for handling various class algorithms or attribute counts.

The possibility of merging two or more algorithms to solve a problem should be researched once you have a better understanding of the advantages and disadvantages of each approach. Utilizing one method's advantages to make up for another's shortcomings is the aim. Finding a single classifier that performs as well as a respectable ensemble of classifiers may be challenging or impossible if we are only concerned with the best classification accuracy achievable..

We think that the difficulties presented by large data will create lots of opportunities for ML. Supervised learning technologies have produced impressive results that were difficult to achieve in previous decades. Supervised learning methods and algorithms can be greatly improved by researchers.

ACKNOWLEDGMENT

All thanks to those who supported me or provided a small contribution to enable me to finish this effort.

REFERENCES

[1] T. M. Mitchell, “Machine learning and data mining,” *Commun. ACM*, vol. 42, no. 11, pp. 30–36, 1999.
[2] D. Chaudhary and E. R. Vasuja, “A review on various algorithms used in machine learning,” *Int. J. Sci. Res. Comput.*

Sci. Eng. Inf. Technol., vol. 5, no. 2, pp. 915–920, 2019. , https://doi.org/10.32628/CSEIT1952248.

[3] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons. b*, vol. 4, pp. 51–62, 2017. doi: 10.20544/HORIZONS.B.04.1.17.P05.

[4] S. Ray, "A quick review of machine learning algorithms," in *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, 2019, pp. 35–39. doi:10.1109/COMITCon.2019.8862451.

[5] A. Géron, *Hands-On Machine Learning with Scikit-Learn & TensorFlow*. 2017.

[6] R. Russell, *Machine Learning: Step-by-Step Guide To Implement Machine Learning Algorithms with Python*. (Knxb), 2020. *http://repository.vnu.edu.vn/handle/VNU_123/92775*.

[7] E. Alpaydin, *Introduction to machine learning*. MIT press, 2020.

[8] J. Bhattacharjee, "Supervised Learning BT - Practical Machine Learning with Rust: Creating Intelligent Applications in Rust," J. Bhattacharjee, Ed. Berkeley, CA: Apress, 2020, pp. 31–105. doi: 10.1007/978-1-4842-5121-8.

[9] I. Muhammad and Z. Yan, "SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY.," *ICTACT J. Soft Comput.*, vol. 5, no. 3, 2015.

[10] G. E. Batista and M. C. Monard, "An analysis of four missing data treatment methods for supervised learning," *Appl. Artif. Intell.*, vol. 17, no. 5–6, pp. 519–533, 2003. https://doi.org/10.1080/713827181.

[11] V. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, 2004. https://doi.org/10.1023/B:AIRE.0000045502.10941.a9.

[12] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, pp. 1205–1224, 2004.

[13] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 161–168.

[14] P. Buczak, H. Huang, B. Forthmann, and P. Doebler, "The machines take over: A comparison of various supervised learning approaches for automated scoring of divergent thinking tasks," *J. Creat. Behav.*, vol. 57, no. 1, pp. 17–36, 2023. **https://doi.org/10.1002/jocb.559**

- [15] P. Chaudhari, H. Agarwal, and V. Bhateja, "Data augmentation for cancer classification in oncogenomics: an improved KNN based approach," *Evol. Intell.*, vol. 14, pp. 489–498, 2021. https://doi.org/10.1007/s12065-019-00283-w

[16] J. Xu, H. Mu, Y. Wang, and F. Huang, "Feature genes selection using supervised locally linear embedding and correlation coefficient for microarray classification," *Comput. Math. Methods Med.*, vol. 2018, 2018.

[17] E. Dritsas and M. Trigka, "Supervised machine learning models for liver disease risk prediction," *Computers*, vol. 12, no. 1, p. 19, 2023. https://doi.org/10.3390/computers12010019

[18] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. data Min. Knowl. Manag. Process*, vol. 5, no. 2, p. 1, 2015. doi : 10.5121/ijdkp.2015.5201.

[19] M. W. Browne, "Cross-validation methods," *J. Math. Psychol.*, vol. 44, no. 1, pp. 108–132, 2000.

[20] P. Geurts, A. Irrthum, and L. Wehenkel, "Supervised learning with decision tree-based methods in computational and systems biology," *Mol. Biosyst.*, vol. 5, no. 12, pp. 1593–1605, 2009. https://doi.org/10.1006/jmps.1999.1279.

[21] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, no. 1, pp. 3–24, 2007.

[22] S. L. Ting, W. H. Ip, and A. H. C. Tsang, "Is Naive Bayes a good classifier for document classification," *Int. J. Softw. Eng. Its Appl.*, vol. 5, no. 3, pp. 37–46, 2011.

[23] A. Aldelemy and R. A. Abd-Alhameed, "Binary Classification of Customer's Online Purchasing Behavior Using Machine Learning," *J. Tech.*, vol. 5, no. 2, pp. 163–186, 2023. https://doi.org/10.51173/jt.v5i2.1226.

[24] AnujSharma and S. Dey, "A comparative study of feature selection and machine learning techniques for sentiment analysis," in *Proceedings of the 2012 ACM research in applied computation symposium*, 2012, pp. 1–7. https://doi.org/10.1145/2401603.2401605.

[25] H. Kang, S. J. Yoo, and D. Han, "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 6000–6010, 2012. https://doi.org/10.1016/j.eswa.2011.11.107.

[26] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proceedings of the 12th international conference on World Wide Web*, 2003, pp. 519–528. https://doi.org/10.1145/775152.775226.

[27] A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble, "Overview of use of decision tree algorithms in machine learning," *Proc. - 2011 IEEE Control Syst. Grad. Res. Colloquium, ICSGRC 2011*, pp. 37–42, 2011. doi:10.1109/ICSGRC.2011.5991826.

[28] A. Bharathi and E. Deepankumar, "Survey on classification techniques in data mining," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 2, no. 7, pp. 1983–1986, E-ISSN: 2347-2693, 2014.
[29] Y.-Y. Song and L. U. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai Arch. psychiatry*, vol. 27, no. 2, p. 130, 2015. doi: 10.11919/j.issn.1002-0829.215044
[30] F. Hajjej, M. A. Alohali, M. Badr, and M. A. Rahman, "A comparison of decision tree algorithms in the assessment of biomedical data," *Biomed Res. Int.*, vol. 2022, 2022. https://doi.org/10.1155/2022/9449497.
[31] M. H. Dunham, *Data mining: Introductory and advanced topics*. Pearson Education India, 2006.
[32] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.
[33] X. Wu and V. Kumar, *The top ten algorithms in data mining*. CRC press, 2009.
[34] S. Suppharangsan, "Comparison and performance enhancement of modern pattern classifiers." University of Southampton, 2010. URI: http://eprints.soton.ac.uk/id/eprint/170393.
[35] Z. G. Hadi, A. R. Ajel, and A. Q. Al-Dujaili, "Comparison Between Convolutional Neural Network CNN and SVM in Skin Cancer Images Recognition," *J. Tech.*, vol. 3, no. 4, pp. 15–22, 2021. https://doi.org/10.51173/jt.v3i4.390.
[36] M. Awad, R. Khanna, M. Awad, and R. Khanna, "Support vector machines for classification," *Effic. Learn. Mach. Theor. Concepts, Appl. Eng. Syst. Des.*, pp. 39–66, 2015.
[37] A. Bilski, "A review of artificial intelligence algorithms in document classification," *Int. J. Electron. Telecommun.*, vol. 57, no. 3, pp. 263–270, 2011. DOI: 10.2478/v10177-011-0035-6
[38] N. R. Dzakiyullah, B. Hussin, C. Saleh, and A. M. Handani, "Comparison neural network and support vector machine for production quantity prediction," *Adv. Sci. Lett.*, vol. 20, no. 10–11, pp. 2129–2133, 2014. https://doi.org/10.1166/asl.2014.5708
[39] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*, vol. 398. John Wiley & Sons, 2013.
[40] M. Srivenkatesh, "Prediction of prostate cancer using machine learning algorithms," *Int. J. Recent Technol. Eng*, vol. 8, no. 5, pp. 5353–5362, 2020.
[41] B. Mahesh, "Machine learning algorithms-a review," *Int. J. Sci. Res. (IJSR).[Internet]*, vol. 9, pp. 381–386, 2020. DOI: 10.21275/ART20203995
[42] S. Cost and S. Salzberg, "A weighted nearest neighbor algorithm for learning with symbolic features," *Mach. Learn.*, vol. 10, pp. 57–78, 1993. https://doi.org/10.1007/BF00993481
[43] S. A. Medjahed, T. A. Saadi, and A. Benyettou, "Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules," *Int. J. Comput. Appl.*, vol. 62, no. 1, 2013. doi: 10.5120/10041-4635.
[44] L. Deng and D. Yu, "Deep learning: methods and applications," *Found. trends® signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2014.
[45] A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo, "Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging," *Cancers (Basel).*, vol. 15, no. 16, p. 4172, 2023.
[46] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. doi:10.1038/nature14539.
[47] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," *J. Pharm. Biomed. Anal.*, vol. 22, no. 5, pp. 717–727, 2000. https://doi.org/10.1016/S0731-7085(99)00272-1.
[48] S.-C. Wang, "Artificial neural network," in *Interdisciplinary computing in java programming*, Springer, 2003, pp. 81–100. https://doi.org/10.1007/978-1-4615-0377-4_5.
[49] A. K. Tyagi, A. Abraham, and A. Kaklauskas, *Intelligent interactive multimedia systems for e-healthcare applications*. Springer, 2022. https://doi.org/10.1007/978-981-16-6542-4.
[50] K. Fukushima, S. Miyake, and T. Ito, "Neocognitron: A neural network model for a mechanism of visual pattern recognition," *IEEE Trans. Syst. Man. Cybern.*, no. 5, pp. 826–834, 1983. doi: 10.1109/TSMC.1983.6313076.
[51] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, 2012. https://doi.org/10.1145/3065386
[52] M. Khan, B. Jan, H. Farman, J. Ahmad, H. Farman, and Z. Jan, "Deep learning methods and applications," *Deep Learn. Converg. to big data Anal.*, pp. 31–42, 2019.
[53] L. R. Medsker, "RECURRENT NEURAL Design and Applications," *New York*, 2001.
[54] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, "Action recognition in video sequences using deep bi-directional LSTM with CNN features," *IEEE access*, vol. 6, pp. 1155–1166, 2017.

[55] T. Jiang, J. L. Gradus, and A. J. Rosellini, "Supervised Machine Learning: A Brief Primer," *Behav. Ther.*, vol. 51, no. 5, pp. 675–687, 2020. doi: 10.1016/j.beth.2020.05.002.

[56] H. Zhang and L. Jiang, "Fine tuning attribute weighted naive Bayes," *Neurocomputing*, vol. 488, pp. 402–411, 2022. https://doi.org/10.1016/j.neucom.2022.03.020.

[57] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artif. Intell. Rev.*, vol. 26, no. 3, pp. 159–190, 2006. https://doi.org/10.1007/s10462-007-9052-3.

[58] R. Kothari and M. Dong, "Decision trees for classification: A review and some new results," *Pattern Recognit. from Class. to Mod. approaches*, pp. 169–184, 2001. https://doi.org/10.1142/9789812386533_0006.

[59] M. O. Khairandish, M. Sharma, V. Jain, J. M. Chatterjee, and N. Z. Jhanjhi, "A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images," *Irbm*, vol. 43, no. 4, pp. 290–299, 2022. https://doi.org/10.1016/j.irbm.2021.06.003.

[60] S. Archana and K. Elangovan, "Survey of classification techniques in data mining," *Int. J. Comput. Sci. Mob. Appl.*, vol. 2, no. 2, pp. 65–71, ISSN: 2321-8363, 2014.

[61] E. Kuzmenko, L. Rumankova, I. Benesova, and L. Smutka, "Czech comparative advantage in agricultural trade with regard to EU-27: Main developmental trends and peculiarities," *Agriculture*, vol. 12, no. 2, p. 217, 2022. https://doi.org/10.3390/agriculture12020217

[62] J. V Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *J. Clin. Epidemiol.*, vol. 49, no. 11, pp. 1225–1231, 1996. https://doi.org/10.1016/S0895-4356(96)00002-9.

[63] M. E. Yahia and B. A. Ibrahim, "K-nearest neighbor and C4. 5 algorithms as data mining methods: advantages and difficulties," *Comput. Syst. Appl.*, vol. 103, pp. 103–109, 2003. doi: 10.1109/AICCSA.2003.1227535.

[64] A. T. Imam, "Relative-fuzzy: A novel approach for handling complex ambiguity for software engineering of data mining models," 2010.

[65] B. Tugrul, E. Elfatimi, and R. Eryigit, "Convolutional neural networks in detection of plant leaf diseases: A review," *Agriculture*, vol. 12, no. 8, p. 1192, 2022. https://doi.org/10.3390/agriculture12081192.