Classification Based on Decision Tree Algorithm for Machine Learning

Abstract

Decision tree classifiers are regarded to be a standout of the most well-known methods to data classification representation of classifiers. Different researchers from various fields and backgrounds have considered the problem of extending a decision tree from available data, such as machine study, pattern recognition, and statistics. In various fields such as medical disease analysis, text classification, user smartphone classification, images, and many more the employment of Decision tree classifiers has been proposed in many ways. This paper provides a detailed approach to the decision trees. Furthermore, paper specifics, such as algorithms/approaches used, datasets, and outcomes achieved, are evaluated and outlined comprehensively. In addition, all of the approaches analyzed were discussed to illustrate the themes of the authors and identify the most accurate classifiers. As a result, the uses of different types of datasets are discussed and their findings are analyzed.

Keywords: *Machine Learning, Supervised, Classification, Decision Tree.*

I. INTRODUCTION

Nowadays, technology has developed a lot, especially in the field of Machine Learning (ML), which is useful for reducing human work. In the field of artificial intelligence, ML integrates statistics and computer science to build algorithms that get more efficient when they are subject to relevant data rather than being given specific instructions. Besides speech recognition, image detection, text localization, etc. ML is the study of computational algorithms that are enhanced from experience automatically. It is considered as an artificial intelligence subset [1, 2]. Orderly to produce foretelling or decision without being specifically programmed to do so, ML algorithms create a model population based on a sample, defined as 'training data' [3, 4]. In a broad area of applications, like email filtering and computer vision, ML algorithms are utilized where it is difficult or impractical to create traditional algorithms to implement functions required [5]. For ML, there are many uses, the most prominent of which is predictive data mining. Two major mechanisms can be broken into ML classification fulfillments; model development and model evaluation [6, 7].

Using the same set of attributes, any instance in every dataset used by ML algorithms is described. The attributes could be continuous, categorical, or binary [8, 9]. If cases are recognized with recognized labels (correct outputs), then learning is called supervised [10, 11]. Supervised Learning is the role of inferring a function from classified training data is machine learning. It also analyzes the testing data and creates a derived task that can be used for new examples to map [12, 13]. Each data input object, however, has a class label pre-assigned. The primary function of supervised algorithms is to learn a model that creates the same labeling preferably for the data offered and popularizes well on unseen data. This is the major aim of algorithms for classification [14].

Classification attempts to predict the goal class with the highest precision. The classification algorithm finds out the relation between the input attribute and the output attribute to construct a model that is a

training process [15 - 17]. The amount of data obtained in data mining environments is huge [18 - 20]. If the data set is properly classified and contains the minimum number of nodes, then using the decision tree method is optimal [21 - 23].

A decision tree is a tree-based technique in which any path beginning from the root is described by a data separating sequence until a Boolean outcome at the leaf node is achieved [24 - 27]. It is the hierarchical exemplification of knowledge relationships that contain nodes and connections. When relations are used to classify, nodes represent purposes [28 - 31].

In this paper, a comprehensive review is performed for the latest and most efficient approaches that have been performed by researchers in the past three years about decision trees in different areas of machine learning. Also, the details of this method, such as using algorithms/approaches, datasets, and the findings achieved are summarized. In addition, this study highlighted the most commonly used approaches and the highest accuracy methods achieved.

The organization of the remaining paper is as follows: Section II contains the decision tree algorithm mentioning its types, benefits, and drawbacks; Section III gives a Literature Review on decision tree Algorithm; Section IV comparison and discussion on the decision tree, and the last section contains the conclusion.

II. Decision Tree Algorithm

One of the widely used techniques in data mining is systems that create classifiers [32]. In data mining, classification algorithms are capable of handling a vast volume of information. It can be used to make assumptions regarding categorical class names, to classify knowledge on the basis of training sets and class labels, and to classify newly obtainable data [33]. Classification algorithms in machine learning contain several algorithms, and in this work, the paper focused on the decision tree algorithm in general. Fig. 1 illustrate a structure of DT.

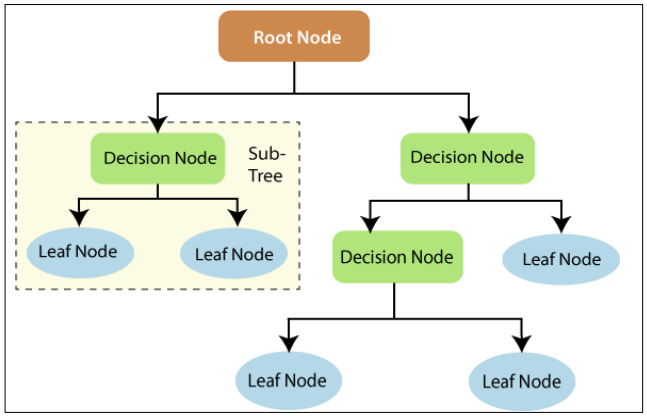

Fig. 1. Decision Tree [34]

Decision trees are one of the powerful methods commonly used in various fields, such as machine learning, image processing, and identification of patterns [35]. DT are a successive model that unites a series of the basic test efficiently and cohesively where a numeric feature is compared to a threshold value in each test [36]. The conceptual rules are much easier to construct than the numerical weights in the neural network of connections between nodes [37, 38]. Mainly for grouping purposes, DT is used. Moreover, DT is a usually utilized classification model in Data Mining [39]. The nodes and branches are composed of each tree. Each node represents features in a category to be classified and each subset defines a value that can be taken by the node [40, 41]. Because of their simple analysis and their precision on multiple data forms, decision trees have found many implementation fields [42]. Fig. 2 show an example of DT.

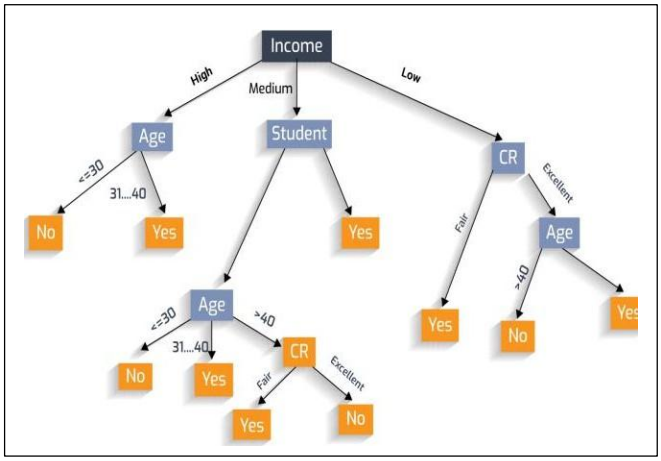

Fig. 2. Example on Decision Tree [43]

A. Types of Decision Tree Algorithms

There are several Types of DT algorithms such as: Iterative Dichotomies 3 (ID3), Successor of ID3 (C4.5), Classification And Regression Tree(CART) [44], CHi-squared Automatic Interaction Detector(CHAID) [45], Multivariate Adaptive Regression Splines (MARS) [46], Generalized, Unbiased, Interaction Detection and Estimation (GUIDE), Conditional Inference Trees (CTREE) [47],[48], Classification Rule with Unbiased Interaction Selection and Estimation (CRUISE), Quick, Unbiased and Efficient Statistical Tree (QUEST) [49], [50]. Table I shown comparison between the frequently used algorithms for the decision tree [51].

B. Entropy and Information Gain

Entropy is employed to measure a dataset's impurity or randomness [52], [53]. The value of entropy always lies between 0 and 1. Its value is better when it is equal to 0 while it is worse when it is equal to 0, i.e. the closer its value to 0 the better. As shown in “Fig. 3”. If the target is G with different attribute values, the entropy of the classification of set S with respect to c states [54], [55]. As shown in “equation (1)”.

$$\text{Entropy}(S) = \sum_{i=1}^{c} P_i \log 2^{P_i} \qquad (1)$$

Where P_i is the ratio of the sample number of the subset and the i-th attribute value.

TABLE I: COMPARISON BETWEEN THE MOST USED ALGORITHMS IN DT

Methods	CART	C4.5	CHAID	QUEST
The measure used for input variable collection	Gini index; Twoing criteria	Entropy info-gain	Chi-square	Chi-square for categorical variables; J-way ANOVA for continuous/ordinal variables
Pruning	Pre-pruning using a single-pass algorithm	Pre-pruning using a single-pass algorithm	Pre-pruning using Chi-square test for independence	Post-pruning
Dependent variable	Categorical/ Continuous	Categorical/ Continuous	Categorical	Categorical
Input variables	Categorical/ Continuous	Categorical/ Continuous	Categorical/ Continuous	Categorical/ Continuous
Split at each node	Binary; Split on linear combinations	Multiple	Multiple	Binary; Split on linear combinations

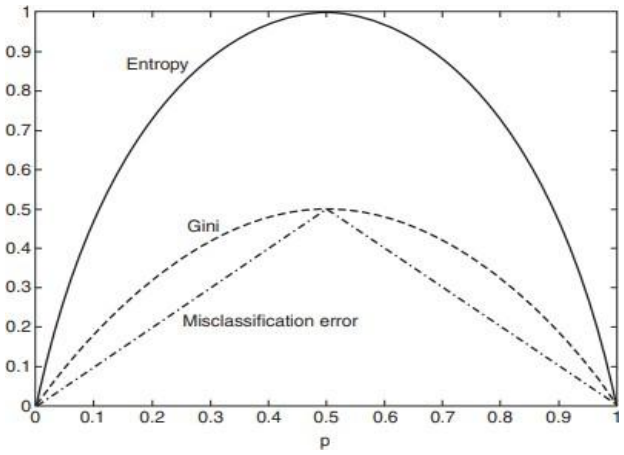

Fig.3. Value of the entropy [56]

Information gain is one metric used for segmentation and is often called mutual information. This intuitively informs how much knowledge of a random variable's value [57, 58]. It's the opposite of entropy, the higher its value is the better. The data gain $Gain(\ S\ ,\ A\ )$ is defined as the following on the definition of entropy [59, 60], as shown in "equation (2)".

$$\text{Gain}(\ \text{S}\ ,\text{A}\ ) = \sum_{\text{v}\ \in\ \text{V(A)}} \frac{|\text{S}_\text{v}|}{|\text{S}|}\ \text{Entropy}(\ \text{S}_\text{v}\ ) \qquad (2)$$

Where the range of attribute A is $V(A)$, and S_v is a subset of set S equal to the attribute value of attribute v [58].

C. Benefits and Drawbacks of decision tree

The DT algorithm is part of the supervised learning algorithm family, and its main objective is to construct a training model that can be used to predict the class or value of target variables through learning decision rules inferred from the training data. The DT algorithm can be used to solve regression and classification problems, but it has benefits and drawbacks [61 - 63], which are summarized in Table II.

TABLE II: BENEFITS AND DRAWBACKS OF DT

Benefits	Drawbacks
1) Simple to comprehend. 2) Quickly translated to a set of principle for production. 3) Can classify both categorical and numerical outcomes, but the attribute generated must be categorical. 4) No a priori hypothesizes are taken with consideration to the goodness of the results.	1) The optimal decision-making mechanism can be deterred and incorrect decisions can follow. 2) There are lots of layers in the decision tree, which makes it interesting. 3) For more training samples, the decision tree's calculation complexity may increase.

III. LITERATURE REVIEW

A decision tree was used in several machine learning and data mining tasks as a classifier. In this study discuss several recent works about the DT. The kinds of Literature Review on DT approaches are summarized in Table III.

Zou et al. [64] Utilized a decision tree (j48), Random Forest (RF), and neural network algorithms for diabetes mellitus prediction. The dataset is physical research data for hospitals in Luzhou, China. There are 14 characteristics involved. Training array randomly extracts data from 68994 stable human and diabetic patients, respectively. They used the full significance of minimum Redundancy Maximum Relevance (mRMR) and Principal Component Analysis (PCA) to minimize dimensionality. In some ways, the effects of RF, as opposed to each other, seemed to be higher than the other classifiers. Also, 0.8084 is the best outcome in the Luzhou data collection.

Assegie and Nair [65], Utilized the DT classification process to classify the handwritten digits of the standard data set of kaggle digits and estimate the accuracy of the model for each digit from 0 to 9. The kaggle features include 42,000 rows and 720 columns used for machine training, vector features are used for pixels of digital images. They used a highly efficient language named "python programming" for the application of machine learning algorithms to map the classifier's success rate graph in the realization of handwritten digits. The findings

suggested that the 83.4% accuracy and decision tree classifier had an impact on handwritten number recognition.

De Felice et al. [66] suggested a decision tree algorithm to recognize known and novel clinical indications before treatment for survival in Locally Advanced Rectal Cancer (LARC). The analytics showed that even non-experts in the field, in particular classification trees, can easily interpret the tree-based machine learning process. Validation errors need to be managed to even achieve their statistical capacity. Around 2007 and 2014, patients with histologically confirmed LARC had their data checked. The Kaplan-Meier approach has been used to determine overall survival (OS). It involved a total of 100 patients. 76.4 % and 71.3% were the 5-year and 7-year OS points. Age, comorbidity, tumor size, Clinical Tumor classification (CT), and clinical node classification are important predictive variables for tree composition (CN). The results showed that the highest survival rates were in elderly patients with a tumor size of less than 5 cm and patients under the age of 65 years who had cT3. A decision tree is a way of getting better clinical practice decision-making, based on broad data sets.

TABLE III: SUMMARY OF LITERATURE REVIEW RELATED OF DT ALGORITHM

Ref.	Year	Dataset	Technique(s)	Accuracy
Nandhini and K.S [75]	2020	UCI	DT , KNN , LR , SVM and NB	DT: 99.93% , KNN: 99.93% , LR: 93.13% , SVM: 90.76% and NB: 79.52%.
Nagra et al. [79]	2020	UCI	SIW-APSO-LS	SIW-APSO-LS: 99.88%
Kuang et al. [71]	2020	SCBs	sSCC	decreasing computational complexity by 47.62% on average
Pathan et al. [78]	2020	images	Optic Disc (OD) segmentation	OD: 99.61%.
Batitis et al. [73]	2020	image	DT	DT: 89.31%
Ramadhan et al. [72]	2020	CICIDS2017	DT and KNN	DT: 99.91% and KNN: 98.94%
Arowolo et al. [77]	2020	RNA-seq Malaria	KNN and DT	KNN: 86.7% and DT: 83.3%.
De Felice et al. [65]	2020	Patients with histologically proven LARC between 2007 and 2014 their data	Kaplan-Meier method	The 5 -year OS rates: 76.4% The 7 -year OS rates:71.3%,
Zhang et al. [74]	2019	Smokers of Chinese Center for Disease Control and Prevention	DT(XGBoost) and RF	DT: 84.11% and RF: 58.11%.
Sathiyanarayanan et al. [82]	2019	Wisconsin Breast Cancer dataset	DT and KNN	DT: 99%. KNN: 97%,
Hu et al. [67]	2019	UCI and COMPAS	OSDT and BinOCT	UCI / OSDT: 66.90% COMPAS/ OSDT: 82.881% and BinOCT: 76.722%
Patil and Kulkarni, [68]	2019	UCI	DST , PT and MLT	DST: The best 99.9% and the worst 81.445%

Sarker et al. [67] presented a Behavioral Decision Tree named "BehavDT" context-aware structure that takes into account consumer behavior-oriented generalization according to the degree of personal choice. In exceptional cases of association, the BehavDT model provided comprehensive decisions as well as context-specific decisions. Experiments were carried out on real smartphone datasets of individual users through the efficiency of the BehavDT model. The results indicated that the Behav DT context-aware model, whose accuracy is up to 90%, is the model that is most energetic compared to other conventional machine learning models.

Hu et al. [68] illustrated the first practical algorithm to optimize decision trees for binary variables. The algorithm is a co-design of analytical limits involving a dedicated bit vector library and data structures that minimize the search area and current application technologies. They used the Binary Optimal Classification Trees (BinOCT) method, which is the current publicly available method, to assess the accuracy and compare it with the Optimal Sparse Decision Trees (OSDT). As well as they utilized text datasets from the University of California, Irvine (UCI) Machine Learning Repository and numeric datasets from the other ProPublica COMPAS datasets. The findings showed that when a COMPAS dataset, the optimal decision tree produced by OSDT, its accuracy 66.90 %. Besides, when BinOCT and OSDT generated the UCI dataset, decision trees, their accuracy is 76.722 %, 82.881 %, respectively.

Patil and Kulkarni [69] introduced a Distributed Spark Tree (DST) to better execute the DT algorithm in terms of model construction time without losing accuracy. Besides, they suggested using them in Spark's climate. Data in Spark's shared architecture does not perform horizontal parallel execution. Spark functions well and coherently in-memory computations, RDD, and map reduction. The dataset that was used from the UCI ML repository and four classes were chosen. Wide data files are utilized to test performance regarding model build time for DST, PySpark (PT), and MLLib (MLT). The findings showed that in terms of accuracy, DST performed better than both PT and MLT, as its lowest value was 81.445 % and the highest according to the scale of the dataset was 99.9 %.

Hussain et al. [70] offered a modern approach, namely a Pixel Label Decision Tree (PLDT), and checks whether it can achieve better detailed femur segmentation efficiency in DXA imaging. PLDT includes extraction and selection of the trait. PLDT was used to uncover secret patterns found in DXA pictures in contrast to photographic images. To decide the best feature set for the model, PLDT generates seven new feature sets and uses Global Threshold (GT), Region Growing Threshold (RGT), and Artificial Neural Networks (ANN). The results revealed that in segmenting DXA images, PLDT

exceeds other conventional partition techniques. For each algorithm such as this PLDT, the accuracy is 91.4%, GT is 68.4%, RGT is 76%, and ANN is 84.4%.

Linty et al. [71] proposed a new approach that affects the amplitude of signals from the Global Navigation Satellite System (GNSS) and was used to detect ionic scintillation events that are concerned with accuracy, reliability, and readiness. A broad collection of 50 Hz post-correlation data was supplied by the GNSS recipient. The outcomes showed that this method, in terms of accuracy and F-score, exceeds state-of-the-art techniques and can achieve a human-driven standard, which is the level of manual annotation. It improves greatly as it gains 98 % of identification, very similar to hand-driven human-driven classification.

Kuang et al. [72] Proposed a structure based on a decision tree named Screen Content Coding (SCC) to make a fast decision in situations by testing their different features in the training sets. Moreover, to prevent the thorough search process, a sequential arrangement of decision trees was illustrated. In addition, SCBs were used as datasets to balance the SCC with the Intra Block Copy (IBC) and PaLeTte (PLT) modes. The results indicated that the SCC system offers a 47.62 % decrease in computational complexity on average, with a small 1.42 % in Bjøntegaard delta bitrate (BDBR).

Ramadhan et al. [73] Demonstrated a comparative analysis of accuracy and process length for each algorithm performed using the K-Nearest Neighbor (KNN) and Decision Tree (DT) algorithms for the detection of DDoS attacks. Moreover, they used the CICIDS2017 dataset that consists of the latest attacks and global packages, is standard and applicable to real-world data in a PCAP format. The findings showed that the accuracy of DT to detect DDoS attacks was higher than the KNN value, the accuracy of DT was 99.91 %, and the accuracy of KNN was 98.94 %.

Batitis et al. [74] presented a system to identify up to 10 irregular red blood cells and to know the accuracy rate for all abnormal red blood cells. Additionally, To detect irregular red blood cells, they employed a DT algorithm in image processing and used frames of former patients for the scheme in hospitals. Also, the camera was used to insert them into the software to capture the slides. The results showed that the accuracy rate averaged 89.31 % and the error rate averaged 10.69 %. Furthermore, the central irregularity of the Codocyte pallor was found to be a cause for the mistake in the classification of abnormal red blood cells.

Zhang et al. [75] Proposed a model based on the decision tree machine learning algorithm named Extreme Gradient Boosting (XGBoost) for the prediction of regular smoking time. Furthermore, to create a simulated data set for smoking time data, the Chinese Center for Disease Control and Prevention collected people's information from smokers. Also, they used a module for extracting feature information. To see its output in the feature extraction module, they used the decision tree (XGBoost) module and Random Forest machine learning algorithms. The results showed that DT efficiency is higher than RF, achieving 84.11 % with DT accuracy, while 58.11 % with RF accuracy.

Nandhini and K.S [76] discussed the effective methods of developing a machine learning model using some of the common algorithms that can distinguish whether mail is spam or ham. UCI's Machine Learning store was used as a dataset for Spambase. Besides, they evaluated the output of Logistic Regression (LR), DT, Naïve Bayes (NB), KNN, and Support Vector Machine (SVM) to construct an efficient machine learning model for spam. Using the Weka tool to train and evaluate the data collection. The results indicated that DT performance is comparable to and better than KNN performance, and the accuracy for both of them is as follows: DT is 99.93 percent, KNN is 99.93 %, LR is 93.13 %, SVM is 90.76 % and NB is 79.52 %.

Taloba and Ismail [77] developed a new machine learning approach for the hybrid decision tree and a genetic algorithm known as GADT for spam detection. The most significant algorithm for enhancing decision tree efficiency is the genetic algorithm. Also, it is efficient and reliable for text classification. A genetic algorithm has used the element of trust that governs decision tree pruning to optimize and detect its optimum value. They used the UCI Machine Learning Store spam dataset. Besides, they used the mechanism of main Principle Component Analysis (PCA) to delete features that are inappropriate for email message content and process them less frequently. The findings showed that after using PCA, the mixed GADT approach has an accuracy of 93.4 % before using PCA and an accuracy of 95.5 %. This implies that the extraction of inappropriate characteristics has a great impact on the PCA.

Arowolo et al. [78] implemented a Principle Component Analysis (PCA) feature extraction algorithm to decrease the dimensions and demonstrate the high dimensions analyze evidence on gene expression. The KNN classification and DT algorithm were utilized to detection various biological structures and to Offer better value resolution as well as to detect new malaria genes and prediction tests. Ribonucleic acid (RNA-seq) sequencing is also used as a data collection. The results indicated that the performance of the KNN classification is better than the DT classification in the PCA feature extraction. The accuracy of KNN reaches 86.7% while the DT reaches 83.3%.

Pathan et al. [79] proposed a new technique that recognizes and removes the blood artery for correct segmentation of the Optic Disc (OD). This is done in two ways. First, the directional filter is used to build an efficient blood vessel identification and exclusion algorithm. In the second step, to detect the contour of the optic disc, the decision tree classifier is utilized to achieve an adaptive threshold. As well as, two separate databases were used, including 300 fundus images obtained from Kasturba Medical College (KMC) Manipal and also the RIM-ONE database that is publicly accessible. The results showed that a fully automatic OD segmentation technique that uses a decision tree classifier to achieve the segmentation threshold improves the robustness of the algorithm even for images containing exudate, vesicle atrophy, and reversals, Hence, resulting in an appropriate fractionation of OD. The mathematical study demonstrates the effect of pretreatment. Therefore, the average values of accuracy

obtained for KMC images are 99.61% and for the RIM-ONE database, the obtained average values of accuracy are 99.15%.

Nagra et al. [80] introduced the Self-Inertia Weight Adaptive Particle Swarm Optimization with Gradient Base Local Search (SIW-APSO-LS) feature selection approach was modified to conduct feature selection and the C4.5 decision tree method was used as a classifier to determine the sub-sets of features given. When comparing algorithms in feature selection problems, 16 datasets from the UCI Machine Learning Repository were used for the experiments. The experimental outcomes demonstrate that SIW-APSO-LS simplifies the collection of features by effectively decreasing the number of features picked, thus maintaining the best precision compared to other literature selection approaches for the same test functions. In the field of attribute collection, the experimental findings showed that the proposed approach is useful and the highest accuracy obtained from a total of 16 datasets is 99.88%

Ahmim et al. [81] proposed a new Intrusion Detection System (IDS) that incorporates diverse classification systems that are DT-based and rule-based concepts, namely the REP tree, JRip algorithm, and Forest PA. In specific, the first and second approaches take data set features as inputs and categorize the network traffic as Attack/Benign. In comparison to the results both the first and second classifiers for reference, the third classifier uses the attributes of the original data collection. The research findings achieved by using the CICIDS2017 dataset to analyze the IDS testify to their dominance in terms of accuracy, identification rate, false alarm rate, and time overhead relative to current state-of-the-art schemes. In thorough, with 94.457%, our model has the highest DR, the highest precision with 96.665%, and the lowest FAR with 1.145%, although its low computing time makes it quickly implemented into a soft real-time system.

Li et al. [82] provided an evidentiary decision tree to classify the fuzzy data set and the ding entropy has been used as an indicator of the partition rules for its construction. Moreover, the Basic Belief Assignment (BBAs) of Iris and wine Datasets are utilized to calculate the optimal splitting feature. The lower the entropy of Deng, the more effective the feature will be to characterize the samples. In contrast to the standard mixture rules employed for the combination of BBAs, the proof DT can be extended specifically to the classification. The findings showed that the implementation of the proof DT based on conviction entropy effectively decreases the complexity of the fuzzy data classification whether the patient is either affected by the cancer type of Malign or Benign. The Wisconsin Breast Cancer dataset, containing 32 attributes and 569 data, was used. They were using a 10 fold cross-validation test to identify and analyze the algorithms. The accuracy is 95% when using Wine datasets, but the accuracy obtained by the Iris datasets is 98%.

Sathiyanarayanan et al. [83] used the DT algorithm under the supervised learning mechanism to reveal breast cancer. Breast cancer identification is conducted here and it focused on data, which separates the data for the preparation and testing process. The result obtained is thus contrasted between the algorithms KNN and DT. The findings reveal that the accuracy obtained by KNN is 97%, while DT reaches the maximum accuracy of 99%. Therefore, a decision tree algorithm that comes under supervised learning methods predicts the type of cancer.

IV. Comparison and Discussion

Decision tree classification algorithms consist of several types that are used to generate DT. This is by the control of both the continuous and periodic attributes of the missing values. DT is generated by a form that is typically represented as a statistical classifier and can be used for clustering. Nodes and branches are included in the DT. Each node requires problems that are based on one or more properties, i.e. comparing an attribute value with a constant or using other functions to compare more than one property. For the purpose of the decision tree, the learning data collection is sometimes referred to as the outcome tree. In order to incorporate classifications in machine learning and data mining using the DT algorithm. In the following sequence, this algorithm is applied iteratively and the classification requires a three-stage process: Construct Model (Learning), Evaluation Model (Accuracy), and Model Use (Classification). The DT classification stage is based on the percentage of acquired information that is measured by entropy. The reach metric is used to describe the test characteristics for a node in the tree and is referred to as the property selection scale (property). As a test function for the current node, the best knowledge property is calculated. Some studies proposed approaches to overcome the shortcomings of the DT problems so that optimal trees can be calculated, based on a review that was performed earlier, without detailed details and samples. DT methods have shown that such problems as described above can be avoided. Furthermore, it will provide the specified dataset with an appropriate solution. According to "Table III" it was observed that in many researches were conducted with different data sets and the DT approach was used to resolve its weaknesses and to obtain better performance. Several optimization techniques have been used in the study [76] to strengthen the decision tree on the UCI ML datasets stored; based on the assessment findings, it was shown that the DT approach got the highest accuracy which is 99.93% comparing to other techniques such as KNN, LR, SVM, and NB which are less performing than the DT approach. In the segmentation task, the study [79] used the DT approach to identify and extract the blood artery for proper Optic Disc (OD) segmentation, which resulted in greater results equal to 99.61%. Moreover, based on the study [69], it has been shown that the DST method can also increase the DT, where both of them used PT and MLT for DT in the UCI datasets; it has been shown that DST is more capable to enhance DT than other techniques. Ultimately, by using UCI Machine Learning Library datasets and CICIDS2017 datasets consisting of the latest attacks among all other datasets, DT proved to be the highest, and their accuracy was the best performing. While the study [75] utilized the DT(XGBoost) and RF on the datasets of the Smokers of the Chinese Center for Disease Control and Prevention, it was found that, again, the DT approach achieved the highest accuracy; which is 84.11%. Furthermore, based on studies [73], [78], [83] using DT and KNN in the CICIDS2017, RNA-seq Malaria, and Wisconsin Breast Cancer datasets, it was found that the DT

approach had the highest accuracy in all three studies. Also, its accuracy was higher when using the CICIDS2017 datasets that achieved 99.91% accuracy.

V. CONCLUSION

Decision tree classifiers are known for their enhanced view of performance outcomes. Because of their strong precision, optimized splitting parameters, and enhanced tree pruning techniques (ID3, C4.5, CART, CHAID, and QUEST) are commonly used by all recognized data classifiers. The separate datasets are used for training samples from a huge data set, which in tum, affects the precision of the test set. Decision trees have several possible concerns about robustness, an adaptation of scalability and optimization of height. But, in contrast to other methods of data classification, decision trees create an efficient rule collection that is simple to understand. This paper reviews the most recent researches that are conducted in many areas, such as analysis of medical diseases, classification of texts, classification of user smartphones and images, etc. Furthermore, the details used in the techniques/algorithms, datasets were used by the authors and achieved outcomes related to the accuracy are summarized for decision trees. Finally, the best accuracy achieved for the decision tree algorithm is 99.93% when it uses a machine learning repository as a dataset.

REFERENCES

[1] D. Abdulqader, A. Mohsin Abdulazeez, and D. Zeebaree, "Machine Learning Supervised Algorithms of Gene Selection: A Review," Apr. 2020.

[2] M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics," *Nature Reviews Genetics*, vol. 16, no. 6, pp. 321–332, 2015.

[3] J. Wang, P. Neskovic, and L. N. Cooper, "Training Data Selection for Support Vector Machines," in *Advances in Natural Computation*, vol. 3610, L. Wang, K. Chen, and Y. S. Ong, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 554–564.

[4] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, 2020.

[5] G. Carleo *et al.*, "Machine learning and the physical sciences," *Reviews of Modern Physics*, vol. 91, no. 4, p. 045002, 2019.

[6] T. Hillel, M. Bierlaire, M. Elshafie, and Y. Jin, "A systematic review of machine learning classification methodologies for modelling passenger mode choice," *Journal of Choice Modelling*, p. 100221, 2020.

[7] D. Zeebaree, H. Haron, A. Mohsin Abdulazeez, and D. Zebari, *Machine learning and Region Growing for Breast Cancer Segmentation*. 2019, p. 93.

[8] C. Feng, S. Wu, and N. Liu, "A user-centric machine learning framework for cyber security operations center," in *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, Beijing, China, Jul. 2017, pp. 173–175, doi: 10.1109/ISI.2017.8004902.

[9] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging artificial intelligence applications in computer engineering*, vol. 160, no. 1, pp. 3–24, 2007.

[10] 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, Nov. 2006, doi: 10.1007/s10462-007-9052-3.

[11] C. Surv, M. N. Murty, P. J. Flynn, A. K. Jain, and P. J. Flynn, *And*. 1999.

[12] D. Sharma and N. Kumar, "A Review on Machine Learning Algorithms, Tasks and Applications," vol. 6, pp. 2278–1323, Oct. 2017.

[13] K. Pahwa and N. Agarwal, "Stock Market Analysis using Supervised Machine Learning," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India, Feb. 2019, pp. 197–200, doi: 10.1109/COMITCon.2019.8862225.

[14] M. Pérez-Ortiz, S. Jiménez-Fernández, P. A. Gutiérrez, E. Alexandre, C. Hervás-Martínez, and S. Salcedo-Sanz, "A Review of Classification Problems and Algorithms in Renewable Energy Applications," *Energies*, vol. 9, no. 8, Art. no. 8, Aug. 2016, doi: 10.3390/en9080607.

[15] Anuradha and G. Gupta, "A self explanatory review of decision tree classifiers," in *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*, Jaipur, India, May 2014, pp. 1–7, doi: 10.1109/ICRAIE.2014.6909245.

[16] S. Patil and U. Kulkarni, "Accuracy Prediction for Distributed Decision Tree using Machine Learning approach," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Apr. 2019, pp. 1365–1371, doi: 10.1109/ICOEI.2019.8862580.

[17] N. S. Ahmed and M. H. Sadiq, "Clarify of the random forest algorithm in an educational field," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, pp. 179–184.

[18] D. Zeebaree, *Gene Selection and Classification of Microarray Data Using Convolutional Neural Network*. 2018.

[19] O. M. Salih Hassan, A. Mohsin Abdulazeez, and V. M. Tiryaki, "Gait-Based Human Gender Classification Using Lifting 5/3 Wavelet and Principal Component Analysis," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, Duhok, Oct. 2018, pp. 173–178, doi: 10.1109/ICOASE.2018.8548909.

[20] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 56–70, 2020.

[21] D. V. Patil and R. S. Bichkar, "A Hybrid Evolutionary Approach To Construct Optimal Decision Trees With Large Data Sets," in *2006 IEEE International Conference on Industrial Technology*, Dec. 2006, pp. 429–433, doi: 10.1109/ICIT.2006.372250.

[22] O. Ahmed and A. Brifcani, "Gene Expression Classification Based on Deep Learning," in *2019 4th Scientific International Conference Najaf (SICN)*, Al-Najef, Iraq, Apr. 2019, pp. 145–149, doi: 10.1109/SICN47020.2019.9019357.

[23] M. A. Sulaiman, "Evaluating Data Mining Classification Methods Performance in Internet of Things Applications," *Journal of Soft Computing and Data Mining*, vol. 1, no. 2, pp. 11–25, 2020.

[24] F. Yang, "An Extended Idea about Decision Trees," in *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, Dec. 2019, pp. 349–354, doi: 10.1109/CSCI49370.2019.00068.

[25] J. Liang, Z. Qin, S. Xiao, L. Ou, and X. Lin, "Efficient and secure decision tree classification for cloud-assisted online diagnosis services," *IEEE Transactions on Dependable and Secure Computing*, 2019.

[26] A. Mohsin Abdulazeez, A. Brifcani, and Issa, "Intrusion Detection and Attack Classifier Based on Three Techniques: A Comparative Study Intrusion Detection and Attack Classifier Based on Three Techniques: A Comparative Study 387," Jan. 2021.

[27] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2670–2679, Apr. 2015, doi: 10.1016/j.eswa.2014.11.009.

[28] A. Shamim, H. Hussain, and Maqbool Uddin Shaikh, "A framework for generation of rules from decision tree and decision table," in *2010 International Conference on Information and Emerging Technologies*, Jun. 2010, pp. 1–6, doi: 10.1109/ICIET.2010.5625700.

[29] A. Suresh, R. Udendhran, and M. Balamurgan, "Hybridized neural network and decision tree based classifier for prognostic decision making in breast cancers," *Soft Computing*, vol. 24, no. 11, pp. 7947–7953, 2020.

[30] Priyanka and D. Kumar, "Decision tree classifier: a detailed survey," *International Journal of Information and Decision Sciences*, vol. 12, no. 3, pp. 246–269, 2020.

[31] A. S. Eesa, A. M. Abdulazeez, and Z. Orman, "A DIDS Based on The Combination of Cuttlefish Algorithm and Decision Tree," *Science Journal of University of Zakho*, vol. 5, no. 4, pp. 313–318, 2017.

[32] R. Kumar and R. Verma, "Classification algorithms for data mining: A survey," *International Journal of Innovations in Engineering and Technology (IJIET)*, vol. 1, no. 2, pp. 7–14, 2012.

[33] S. S. Nikam, "A comparative study of classification techniques in data mining algorithms," *Oriental journal of computer science & technology*, vol. 8, no. 1, pp. 13–19, 2015.

[34] C. Z. Janikow, "Fuzzy decision trees: issues and methods," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 28, no. 1, pp. 1–14, 1998.

[35] G. Stein, B. Chen, A. S. Wu, and K. A. Hua, "Decision tree classifier for network intrusion detection with GA-based feature selection," in *Proceedings of the 43rd annual Southeast regional conference-Volume 2*, 2005, pp. 136–141.

[36] I. S. Damanik, A. P. Windarto, A. Wanto, S. R. Andani, and W. Saputra, "Decision Tree Optimization in C4. 5 Algorithm Using Genetic Algorithm," in *Journal of Physics: Conference Series*, 2019, vol. 1255, no. 1, p. 012012.

[37] R. Barros, M. Basgalupp, A. de Carvalho, and A. Freitas, "A Survey of Evolutionary Algorithms for Decision-Tree Induction," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 42, pp. 291–312, Jan. 2012, doi: 10.1109/TSMCC.2011.2157494.

[38] G. Gupta, "A self explanatory review of decision tree classifiers," in *International conference on recent advances and innovations in engineering (ICRAIE-2014)*, 2014, pp. 1–7.

[39] S. S. Gavankar and S. D. Sawarkar, "Eager decision tree," in *2017 2nd International Conference for Convergence in Technology (I2CT)*, Mumbai, Apr. 2017, pp. 837–840, doi: 10.1109/I2CT.2017.8226246.

[40] P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," *IEEE Transactions on Geoscience Electronics*, vol. 15, no. 3, pp. 142–147, 1977.

[41] A. Dey, "Machine learning algorithms: a review," *International Journal of Computer Science and Information Technologies*, vol. 7, no. 3, pp. 1174–1179, 2016.

[42] J. Mrva, Š. Neupauer, L. Hudec, J. Ševcech, and P. Kapec, "Decision Support in Medical Data Using 3D Decision Tree Visualisation," in *2019 E-Health and Bioengineering Conference (EHB)*, Nov. 2019, pp. 1–4, doi: 10.1109/EHB47216.2019.8969926.

[43] Y. Bengio, O. Delalleau, and C. Simard, "DECISION TREES DO NOT GENERALIZE TO NEW VARIATIONS," *COMPUTATIONAL INTELLIGENCE*, p. 19.

[44] C. E. Brodley and P. E. Utgoff, "Multivariate decision trees," *Machine learning*, vol. 19, no. 1, pp. 45–77, 1995.

[45] G. K. F. Tso and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, no. 9, pp. 1761–1768, Sep. 2007, doi: 10.1016/j.energy.2006.11.010.

[46] S. Singh and P. Gupta, "Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey," *International Journal of Advanced Information Science and Technology (IJAIST)*, vol. 27, no. 27, pp. 97–103, 2014.

[47] L. Rokach and O. Maimon, "Top-Down Induction of Decision Trees Classifiers—A Survey," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 35, pp. 476–487, Dec. 2005, doi: 10.1109/TSMCC.2004.843247.

[48] T.-S. Lim, W.-Y. Loh, and Y.-S. Shih, "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms," *Machine learning*, vol. 40, no. 3, pp. 203–228, 2000.

[49] W.-Y. Loh, "Fifty Years of Classification and Regression Trees," *International Statistical Review*, vol. 82, Jun. 2014, doi: 10.1111/insr.12016.

[50] S. R. Jiao, J. Song, and B. Liu, "A Review of Decision Tree Classification Algorithms for Continuous Variables," in *Journal of Physics: Conference Series*, 2020, vol. 1651, no. 1, p. 012083.

[51] Y.-Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, pp. 130–5, Apr. 2015, doi: 10.11919/j.issn.1002-0829.215044.

[52] RekhaMolala, "Entropy, Information gain and Gini Index; the crux of a Decision Tree," *Medium*, Mar. 23, 2020. https://blog.clairvoyantsoft.com/entropy-information-gain-and-gini-index-the-crux-of-a-decision-tree-99d0cdc699f4 (accessed Dec. 28, 2020).

[53] V. Cheushev, D. A. Simovici, V. Shmerko, and S. Yanushkevich, "Functional entropy and decision trees," in *Proceedings. 1998 28th IEEE International Symposium on Multiple-Valued Logic (Cat. No. 98CB36138)*, 1998, pp. 257–262.

[54] X. Chen, Z. Yang, and W. Lou, "Fault Diagnosis of Rolling Bearing Based on the Permutation Entropy of VMD and Decision Tree," in *2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)*, Xiamen, China, Oct. 2019, pp. 1911–1915, doi: 10.1109/EITCE47263.2019.9095187.

[55] C. Shang, M. Li, S. Feng, Q. Jiang, and J. Fan, "Feature selection via maximizing global information gain for text classification," *Knowledge-Based Systems*, vol. 54, pp. 298–309, Dec. 2013, doi: 10.1016/j.knosys.2013.09.019.

[56] T. Maszczyk and W. Duch, "Comparison of Shannon, Renyi and Tsallis entropy used in decision trees," in *International Conference on Artificial Intelligence and Soft Computing*, 2008, pp. 643–651.

[57] L. E. Raileanu and K. Stoffel, "Theoretical Comparison between the Gini Index and Information Gain Criteria," *Annals of Mathematics and Artificial Intelligence*, vol. 41, no. 1, pp. 77–93, May 2004, doi: 10.1023/B:AMAI.0000018580 96245.c6.

[58] Y. Liu, L. Hu, F. Yan, and B. Zhang, "Information Gain with Weight Based Decision Tree for the Employment Forecasting of Undergraduates," in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, Beijing, China, Aug. 2013, pp. 2210–2213, doi: 10.1109/GreenCom-iThings-CPSCom.2013.417.

[59] R. L. De Mántaras, "A distance-based attribute selection measure for decision tree induction," *Machine learning*, vol. 6, no. 1, pp. 81–92, 1991.

[60] S. Taneja, C. Gupta, K. Goyal, and D. Gureja, "An enhanced k-nearest neighbor algorithm using information gain and clustering," in *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, 2014, pp. 325–329.

[61] Y. Zhao and Y. Zhang, "Comparison of decision tree methods for finding active objects," *Advances in Space Research*, vol. 41, no. 12, pp. 1955–1959, 2008.

[62] K. Mittal, D. Khanduja, and P. C. Tewari, "An insight into 'Decision Tree Analysis'", *World Wide Journal of Multidisciplinary Research and Development*, vol. 3, no. 12, pp. 111–115, 2017.

[63] Priyanka and D. Kumar, "Decision tree classifier: a detailed survey," *International Journal of Information and Decision Sciences*, vol. 12, no. 3, pp. 246–269, 2020.

[64] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in genetics*, vol. 9, p. 515, 2018.

[65] T. A. Assegie and P. S. Nair, "Handwritten digits recognition with decision tree classification: a machine learning approach," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 5, p. 4446, 2019.

[66] F. De Felice *et al.*, "Decision tree algorithm in locally advanced rectal cancer: an example of over-interpretation and misuse of a machine learning approach," *Journal of Cancer Research and Clinical Oncology*, vol. 146, no. 3, pp. 761–765, 2020.

[67] I. H. Sarker, A. Colman, J. Han, A. I. Khan, Y. B. Abushark, and K. Salah, "Behavdt: a behavioral decision tree learning to build user-centric context-aware predictive model," *Mobile Networks and Applications*, vol. 25, no. 3, pp. 1151–1161, 2020.

[68] X. Hu, C. Rudin, and M. Seltzer, "Optimal sparse decision trees," in *Advances in Neural Information Processing Systems*, 2019, pp. 7267–7275.

[69] S. Patil and U. Kulkarni, "Accuracy Prediction for Distributed Decision Tree using Machine Learning approach," in *2019 3rd International Conference on Trends in Electronics and Informatics*

(ICOEI), Apr. 2019, pp. 1365–1371, doi: 10.1109/ICOEI.2019.8862580.

[70] D. Hussain, M. A. Al-Antari, M. A. Al-Masni, S.-M. Han, and T.-S. Kim, "Femur segmentation in DXA imaging using a machine learning decision tree," *Journal of X-ray Science and Technology*, vol. 26, no. 5, pp. 727–746, 2018.

[71] N. Linty, A. Farasin, A. Favenza, and F. Dovis, "Detection of GNSS Ionospheric Scintillations Based on Machine Learning Decision Tree," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 1, pp. 303–317, Feb. 2019, doi: 10.1109/TAES.2018.2850385.

[72] W. Kuang, Y. Chan, S. Tsang, and W. Siu, "Machine Learning-Based Fast Intra Mode Decision for HEVC Screen Content Coding via Decision Trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 5, pp. 1481–1496, May 2020, doi: 10.1109/TCSVT.2019.2903547.

[73] I. Ramadhan, P. Sukarno, and M. A. Nugroho, "Comparative Analysis of K-Nearest Neighbor and Decision Tree in Detecting Distributed Denial of Service," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, Yogyakarta, Indonesia, Jun. 2020, pp. 1–4, doi: 10.1109/ICoICT49345.2020.9166380.

[74] V. M. E. Batitis, M. J. G. Caballes, A. A. Ciudad, M. D. Diaz, R. D. Flores, and E. R. E. Tolentin, "Image Classification of Abnormal Red Blood Cells Using Decision Tree Algorithm," in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, Mar. 2020, pp. 498–504, doi: 10.1109/ICCMC48092.2020.ICCMC-00093.

[75] Y. Zhang, J. Liu, Z. Zhang, and J. Huang, "Prediction of Daily Smoking Behavior Based on Decision Tree Machine Learning Algorithm," in *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, Jul. 2019, pp. 330–333, doi: 10.1109/ICEIEC.2019.8784698.

[76] S. Nandhini and J. M. K.S, "Performance Evaluation of Machine Learning Algorithms for Email Spam Detection," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, Feb. 2020, pp. 1–4, doi: 10.1109/ic-ETITE47903.2020.312.

[77] A. I. Taloba and S. S. I. Ismail, "An Intelligent Hybrid Technique of Decision Tree and Genetic Algorithm for E-Mail Spam Detection," in *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, Dec. 2019, pp. 99–104, doi: 10.1109/ICICIS46948.2019.9014756.

[78] M. O. Arowolo, M. Adebiyi, A. Adebiyi, and O. Okesola, "PCA Model For RNA-Seq Malaria Vector Data Classification Using KNN And Decision Tree Algorithm," in *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, Mar. 2020, pp. 1–8, doi: 10.1109/ICMCECS47690.2020.240881.

[79] S. Pathan, P. Kumar, R. Pai, and S. V. Bhandary, "Automated detection of optic disc contours in fundus images using decision tree classifier," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 52–64, 2020.

[80] A. A. Nagra *et al.*, "Hybrid self-inertia weight adaptive particle swarm optimisation with local search using C4. 5 decision tree classifier for feature selection problems," *Connection Science*, vol. 32, no. 1, pp. 16–36, 2020.

[81] A. Ahmim, L. Maglaras, M. A. Ferrag, M. Derdour, and H. Janicke, "A novel hierarchical intrusion detection system based on decision tree and rules-based models," in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2019, pp. 228–233.

[82] M. Li, H. Xu, and Y. Deng, "Evidential decision tree based on belief entropy," *Entropy*, vol. 21, no. 9, p. 897, 2019.

[83] P. Sathiyanarayanan, S. Pavithra, M. S. SARANYA, and M. Makeswari, "Identification of Breast Cancer Using The Decision Tree Algorithm," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019, pp. 1–6.