Journal of Applied Information Science
8 (1&2), 2020, 14-21
http://www.publishingindia.com/jais

Comparison and Analysis of Machine Learning Techniques for the Prediction of Acute Appendicitis

Rijuta Goswami[1*], Haneet Kour[2], Jatinder Manhas[3], Vinod Sharma[4]
[1]M.Tech. Student, Department of Computer Science and IT, University of Jammu, Jammu and Kashmir, India. Email: rijuta2508@gmail.com
[2]PhD Research Scholar, Department of Computer Science and IT, University of Jammu, Jammu and Kashmir, India. Email: haneetkour9@gmail.com
[3]Sr. Assistant Professor, Department of Computer Science and IT, Bhaderwah Campus, University of Jammu, Jammu and Kashmir, India. Email: manhas.jatinder@gmail.com
[4]Professor, Department of Computer Science and IT, University of Jammu, Jammu and Kashmir, India. Email: vnodshrma@gmail.com
*Corresponding Author

Abstract: Appendicitis is the most serious medical emergency requiring surgery for removing the appendix. Appendicitis treatment needs physical examination accompanied by blood tests and imaging scans to better detect signs of appendicitis or to rule out potential causes of the symptoms. Diagnosing appendicitis can be difficult because of the proximity of the appendix to other pelvic organs and its location, thus its symptoms have a tendency to overlap with other illnesses. The aim of the current study is to compare and analyze the performance of machine learning (ML) techniques in the prediction of appendicitis accurately. In the current paper, three machine learning techniques namely Support Vector Machine (SVM), Decision Tree and K-nearest Neighbor (KNN) have been taken. The experiments were carried out on the benchmark dataset of Appendicitis consisting of 590 patients. The performance of these ML techniques has been evaluated on the basis of three measures i.e. Accuracy, Recall, and Precision. The experimental result revealed that the Decision Tree algorithm performed better with an accuracy of 73.72%, Precision of 75.35%, and Recall of 68.64% as compared to SVM and KNN. It can be inferred from the experimental results that models based on machine learning techniques can predict appendicitis accurately and can serve as a decision-making aid by providing a correct and timely diagnosis of appendicitis, thereby reducing the negative appendectomy rate.

Keywords: Appendicitis, Decision tree, K-Nearest neighbor, Machine learning, Support vector machine.

I. Introduction

The appendix is a pouch-like structure which is attached to the large intestine at the beginning and has no defined function. A condition in which the appendix becomes inflamed and filled with pus causing pain is called appendicitis. It occurs most frequently with a mean age of 28 between the ages of 5 and 45. Males are significantly more predisposed to experience acute appendicitis than females, with a lifetime incidence of 8.6% for males and 6.7% for females. There are about 300,000 hospital visits in the United States annually for appendicitis-related problems [1]. This is a medical emergency that involves surgery, called appendectomy, as soon as possible.

There is no single test available to diagnose appendicitis based on the symptoms. If appendicitis is suspected, the doctor will check for the intensity of pain, body temperature, tenderness in the right iliac fossa (RIF) and rebound. Depending on the results of the physical examination, the doctor may order for laboratory tests which include measuring blood cell counts, CRP and imaging tests like CT scans and ultrasounds check for the signs of appendicitis and to rule out other illnesses. Due to the proximity of appendix with other pelvic organs and its varying size and location, the probability of misdiagnosis is quite high. These tests are helpful if the diagnosis is unclear. Although beneficial, there are health threats associated with them as they require exposure to harmful radiations and a large section of society may not be able to afford these expensive tests.

The healthcare industry is considered to be highly suitable for the applications of Machine learning and AI tools. In recent years machine learning has enhanced the quality of automation and intelligent decision making [2]. Kothainayaki and Thangaraj (2013) [3] used the K-means algorithm to classify diabetic dataset. The k-means algorithm is well known for its efficiency in clustering large data sets. In the same year, Gupta (2013) [4] developed a model for mining fuzzy amino acid associations in peptide sequences of the herpes virus. Manonmani *et al*. (2018)

[5] worked on the conversion of 2D medical imaging data to 3D images using Laplacian process for arriving at the required result with better efficiency. Hashim *et al.* (2019) [6] developed a mobile application that enables users to diagnose medical conditions based on the symptoms provided by the users using Deep Learning. Like these, a wide variety of exciting and future-looking applications of AI/ML techniques have benefitted the space of healthcare. The goal of the current research is to integrate healthcare technology and use machine learning techniques to predict appendicitis accurately in suspected patients. In the current paper, the authors evaluated the performance of machine learning techniques in order to find out the most accurate ML technique for the prediction of Appendicitis.

This research paper is structured as follows: Section II presents Literature Review, Section III discusses Materials and Methodology adopted to carry out the experiments, Section IV explains Results and Discussion, and Section V describes Conclusion and Future Scope.

II. Literature Review

In the recent past, significant research has been done in the domain of health care using machine learning techniques to predict acute appendicitis. A study done by various researchers for prediction of acute appendicitis using machine learning techniques during the past few years is presented as under:

Prabhudesai *et al. (2007)* [7] examined the role of Artificial Neural Network (ANN's) in the diagnosis of appendicitis in patients with acute right iliac fossa (RIF) pain and compared its performance with the assessment made by clinicians using Alvarado Scoring System (ASS). 60 consecutive patients presented with suspected appendicitis in a teaching hospital over 6 months were included in this study. The NTS Weightless Neural System (Novel Technical Solutions, London, UK), a multilayered perceptron-type ANN was used in this study. The experimental results showed that ANNs can accurately diagnose appendicitis. The sensitivity, specificity, and positive and negative predictive values of the ANN were 100%, 97.2%, 96.0%, and 100% respectively.

After two years, Sivsankar *et al.* (2009) [8] implemented Bayesian classifier and Backpropagation Neural Network classifier for diagnosis of appendicitis in patients with RIF pain using Alvarado Scoring System (ASS) by employing the dataset of 2230 records collected from BHEL Hospital, Tiruchirappalli, India. Input Parameters consists of pain site, pain nature, nausea, previous surgery and the output parameters are different classes of appendicitis namely mild (Inflammation only), moderate (Inflammation, Faceolith and Turgid) and severe (Gangrenous and Perforated) appendicitis.

Hsien-Wei *et al.* (2010) [9] presented a model based on Decision Tree for the treatment of acute appendicitis by employing the dataset of 532 patients. The decision tree algorithm was constructed with the data mining workbench Clementine version 8.1. The Standard *t*-test and $\chi 2$-*test* were used for statistical analysis. The new model was more convenient and the accuracy of the diagnostic rate of the new model was better than ASS. The sensitivity and specificity of the presented model were found to be 0.945 and 0.805, respectively.

Next year, O. Yoldas *et al.* (2011) [10] implemented a neural network for the diagnosis of acute appendicitis using a dataset of 156 patients' records. A three-layered perceptron ANN model has been developed with the backpropagation algorithm on SPSS 19 platform. The sensitivity, specificity, and positive and negative predictive values of the ANN were found to be 100%, 97.2%, 88.0%, and 100%, respectively.

In the same year, Ohle *et al.* (2011) [11] presented a study on The Alvarado Clinical Scoring System (ACSS) to stratify patients with symptoms of suspected appendicitis. The aim was to assess the diagnostic accuracy of the Alvarado score. The diagnostic accuracy of the score was analyzed at the two cut-off points: score of 5 (1 to 4 vs. 5 to 10) and score of 7 (1 to 6 vs. 7 to 10). In terms of diagnostic accuracy, the cut-point of 5 was good at 'ruling out' admission for appendicitis (sensitivity 99% overall, 96% men, 99% woman, 99% children). At the cut-point of 7, recommended for 'ruling in' appendicitis and progression to surgery, the score performed poorly in each subgroup (specificity overall 81%, men 57%, woman 73%, children 76%). From the results of the study, it can be concluded that the Alvarado score is a useful diagnostic 'rule out' score at a cut point of 5 for all patient groups.

Balu and Devi (2012) [12] used sonographic images as input and developed an image mining system to automate the diagnosis of acute appendicitis with a significant reduction in time. Real dataset of 44 patient's sonographic image was collected. Region-based segmentation algorithm followed by Euclidean distance method yielded an accurate diagnosis of appendicitis with better accuracy, greater noise reduction, faster speed and greater automation and yielded a sensitivity of 86% and specificity of 81%.

Jade *et al.* (2016) [13] conducted a study to evaluate Modified Alvarado Scoring System for the diagnosis of acute appendicitis and its correlation by histopathology. Pre-operatively Modified Alvarado Score was assigned to all patients and the results were compared with operative and histopathological diagnosis reports. In this study, Modified Alvarado score of 1-4, 5-7 and 8-10 had the accuracy of 10%, 75% and 100% respectively. Higher the score higher was the accuracy. The experimental results predicted the score is more sensitive to males' patients and it is reliable.

III. Materials and Methods

For the current research, Appendicitis dataset was collected online [14] (https://osf.io/9wvys), and originally from the Department of Pediatric Surgery of a hospital in Berlin. The dataset had 590 records of patients falling between the age group 1-17. All these patients had undergone appendectomies.

A total of 12 features were selected which are summarized in Table I. The parameters taken into consideration were mostly laboratory test values and the sonography report, which was used to determine the appendix diameter. The experiments were carried out on Matlab platform. The aim of the study research was to predict appendicitis in suspected patients as well as to differentiate between complicated cases i.e. patients with perforated appendix and uncomplicated cases i.e. patients with infected appendix and formation of an abscess, in order to determine which cases would require immediate treatment. The output variable was multiclass i.e. there are three class labels: Complicated (1) - those cases which require immediate treatment, Uncomplicated (2) - cases with the need to kept under observation to gain more clarity about the diagnosis and Negative (3) - the cases which did not have appendicitis. The overall methodology of the current study is depicted in Fig. 1.

Table I: Parameters in the Dataset with Mean Value for Each Class (N = 590)

	Complicated (1)	Uncomplicated (2)	Negative (3)
N (%)	183 (31.01%)	290 (49.15%)	117 (19.8%)
No. of Males (%)	112 (61.02%)	164 (56.55%)	48 (41.02%)
No. of Females (%)	71 (38.79%)	126 (43.44%)	69 (58.97%)
Average Age (in years)	9.43	10.5	12
Average Age of Males (in years)	9.58	10.35	9.66
Average Age of Females (in years)	9.19	10.86	13.63
CRP (mg/L)	67.82	19.60	18.997
Thrombocytes (x 10 9/L)	291.87	276.76	266.17
Leukocytes (x 10 9/L)	15.72	13.86	10.2
Neutrophils (x 10 9/L)	13.03	11.01	7.05
Immature Granulocytes (x 10 9/L)	0.15	0.062	0.032
Lymphocytes (x 10 9/L)	1.45	1.79	2.13
Monocytes (x 10 9/L)	1.15	0.92	0.77
Eosinophils (x 10 9/L)	0.044	0.11	0.164
Basophils (x 10 9/L)	0.027	0.031	0.031
Appendiceal Diameter (mm)	10.45	9.32	7.16

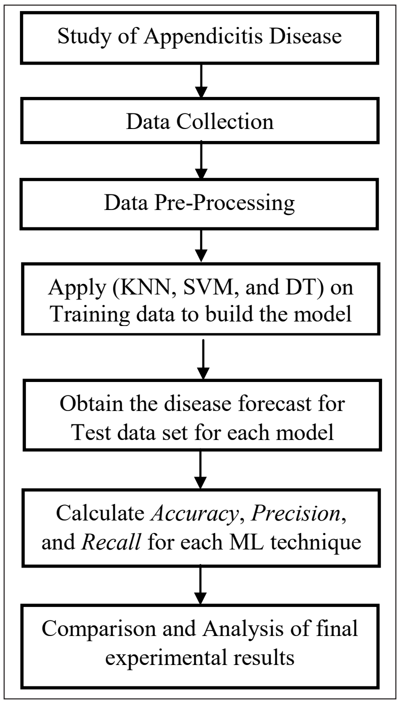

Fig. 1. Flow Chart of the Current Study

After data collection, data pre-processing has been done in order to convert nominal values into numerical values, to detect outliers using kurtosis, and to perform imputation using statistical techniques. The dataset was a mixture of continuous as well as categorical variables originally hence all the variables were converted to numerical form. The features in the dataset were on different scales so the dataset was normalized. After this, the dataset is divided into two sets in the ratio 80: 20 i.e. *training set* with 472 instances and the *testing set* with 118 instances. Three machine learning algorithms namely Support Vector Machine (SVM), Decision Tree and K- nearest Neighbor (KNN) were implemented on the training set to build the model for Appendicitis prediction. The performance of each trained model has been evaluated on the test data by computing three performance measures such as Accuracy, Recall and Precision.

A. Implementation using K-Nearest Neighbor

It is one of the essential classification algorithms in Machine Learning and belongs to the supervised learning domain. Given the training dataset, it classifies coordinates into groups identified by an attribute and given another set of data points called testing dataset, it allocates these points a group by analyzing the training set. It is a model that classifies data points based on the points that are most similar to it [15].

In the current research work, the KNN classifier was first trained using a dataset with predictor variables (x1, x2, x3, …., x12). There is no structured method available to find the best value for 'k'. Choosing a smaller value 'k' can be noisy whereas choosing a larger value of 'k' can make the boundaries between classes less distinct. After a series of trial and error, the value of K was chosen as K = 25 (described by NumNeighbors property in Matlab). Various distance metrices can be used to calculate the distance between test data and each row of the training dataset. The allowable distance metric for KNN is cityblock, chebychev, cosine, Euclidean, mahalanobis, minkowski etc. Euclidean distance measure was used to calculate the distance between the datapoints for the current study. Trained *k*-nearest neighbor classification model, returned as a Classification KNN model object. The values of the class labels for test instances were predicted based on largest posterior probabilities. Fig. 2 describes the KNN model developed in Matlab.

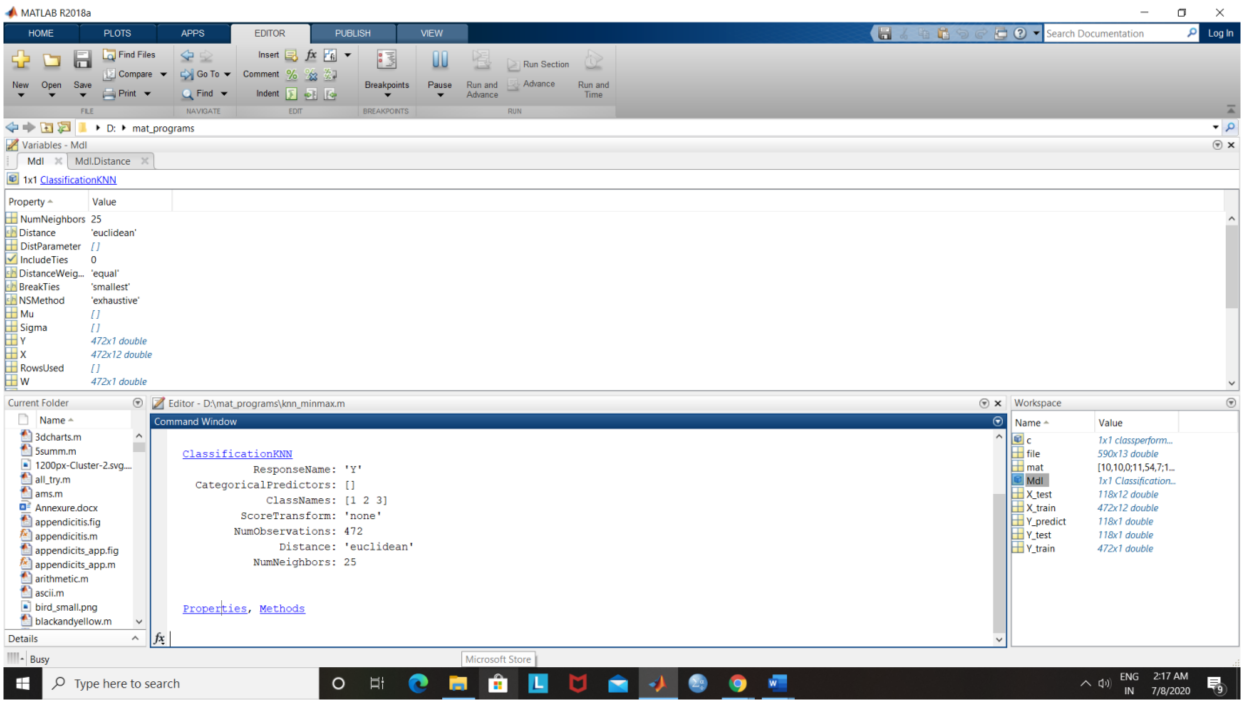

Fig. 2: Implementation of KNN Classifier in Matlab

B. Implementation using Support Vector Machine (SVM)

SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. The objective SVM model is to find the best line in two dimensions or the best hyperplane in more than two dimensions in order to help us separate our space into classes. The hyperplane (line) is found through the *maximum margin*, i.e., the maximum distance between data points of both classes [16].

For implementing SVM classifier model in the current multi-class classification problem, Error-Correcting Output Codes (ECOC) was used. ECOC is basically an ensemble model that implements multiple binary learners to solve the multi-class classification problem. In Matlab the ECOC model uses SVM binary learners by default. In the current model, three binary Support Vector machines were used. Numerous coding designs are available to choose in Matlab like one vs one, binary complete, one vs all etc. A one-versus-one coding design for three SVM learners was used for the current model. The number of binary learners for this coding design is calculated using the formula "k(k-1)/2" where k = number of classes. In our case, the value of k = 3, hence three classes yield three binary learners. Fig. 3 represents the SVM model and the coding matrix of the model based on one vs one coding. The columns of the coding matrix correspond to the learners, and the rows correspond to the classes. In this coding design, for each binary learner, one class is positive, another is negative, and the software ignores the rest. This design exhausts all combinations of class pair assignments. Kernel function used was a linear kernel.

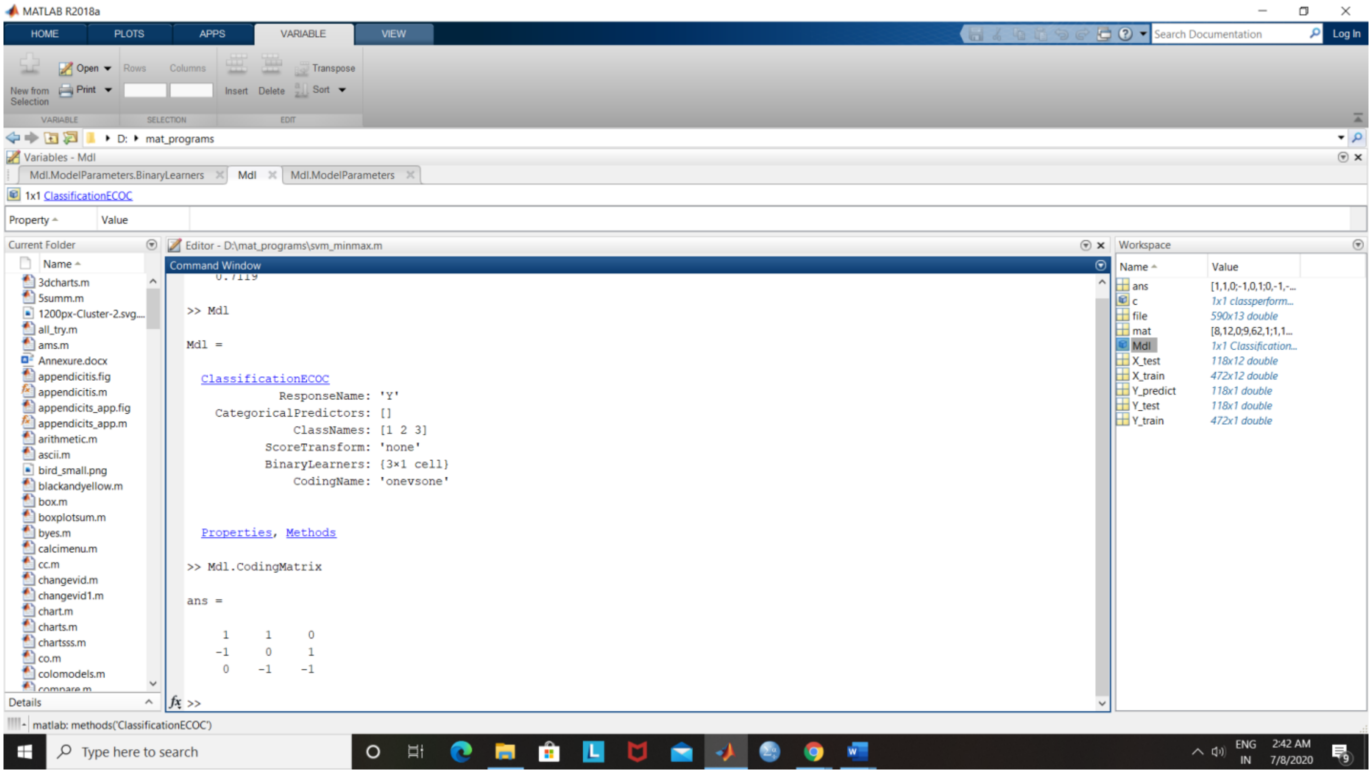

Fig. 3: Implementation of SVM Classifier in Matlab

C. Implementation using Decision Tree

Decision tree algorithm falls under the category of supervised learning. They can be used to solve both regression and classification problems. It uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree [17].

For the current research MATLAB platform was used to fit a binary decision tree for the multiclass classification problem based on the training dataset with 12 predictor variables (x1, x2, x3, …., x12). Matlab grows deep decision trees by default. The default values of the tree depth controller, MaxNumSplits is n-1, where n is the training sample size. These default values tend to grow deep trees for large training sample sizes. For our study the training sample size n would be 472, thereby leading to a very deep binary tree. Hence the depth of the tree was controlled by restricting the maximum number of splits to 6. Shallower trees to reduce model complexity or computation time. The resultant binary tree had 13 nodes with 7 leaf nodes and 6 internal nodes. Fig. 3 represents the binary decision tree model's graphic representation. The algorithm used to select the best split predictor at each node was 'allsplits' which selects the split predictor that maximizes the split-criterion gain over all possible splits of all predictors. The split criterion used at each node was Gini Index. The attributes selected by the split criterion are x7 = immature granulocytes, x12 = appendiceal diameter, x11 = basophils, x3 = CRP.

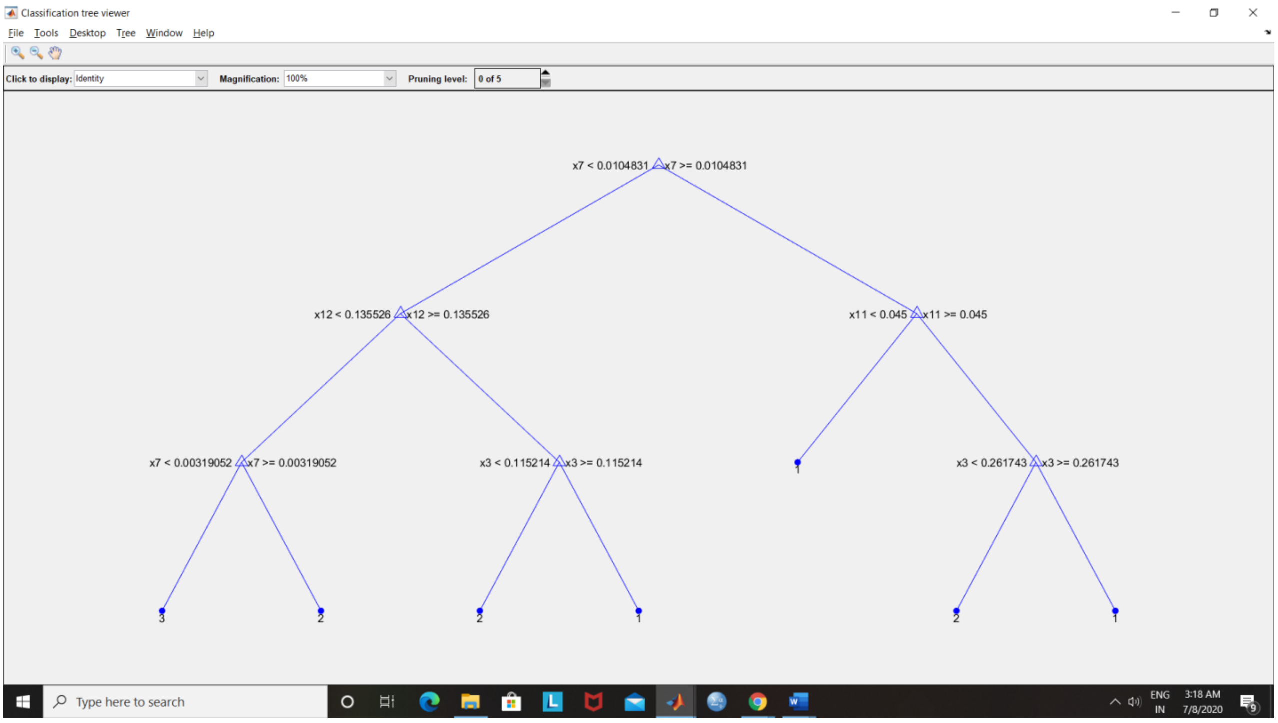

Fig. 4: Implementation of Decision Tree Classifier in Matlab

IV. Result and Analysis

In this research study, a total of 590 records of patients along 12 features related to acute appendicitis were undertaken. The male to female ratio was 162:133 (324 males and 266 females). All the patients had undergone surgery, but the histopathological reports were later used to determine how many actually had appendicitis and how many did not. Through histopathology it was found that out of 590 cases, 183 (31%) cases were complicated cases of acute appendicitis, 290 (49.15%) were uncomplicated cases and 117 (19.8%) cases were those who had undergone negative appendectomies. After the models were trained, each model predicted whether the case of appendicitis was complicated, uncomplicated or negative for the instances in the testing dataset. The performance of each model was evaluated on the basis of 3 parameters: Accuracy, Precision, and Recall.

From the experimental results, it has been found out that all the machine learning techniques can be used for Appendicitis prediction. The experimental results for KNN, SVM and Decision Tree technique are presented in Table II, Table III and Table IV respectively. Table V presents the overall results for all three techniques. KNN model predicted accuracy of 61.02%, recall of 45% and precision of 53% on un-normalized data. With normalization, the performance of KNN has been improved as it predicted accuracy of 71.18%, recall of 63% and a precision of 68.6% on normalized data. SVM model predicted accuracy of 71.19%, recall of 68% and precision of 67% on un-normalized data; and accuracy of 71.19%, recall of 65.09% and precision of 70.06% on normalized data. Decision tree model predicted accuracy of 72.88%, recall of 68.18% and precision of 71% on un-normalized data; and accuracy of 73.72%, recall of 68.64% and precision of 75.35% on normalized data.

Table II: Experimental Results for KNN Model

Parameter	Without Normalization	With Normalization
Accuracy	61.02%	71.18%
Recall	45%	63%
Precision	53%	68.6%

Table III: Experimental Results for SVM Model

Parameter	Without Normalization	With Normalization
Accuracy	71.19%	71.19%
Recall	68%	65.09%
Precision	67%	70.06%

Table IV: Experimental Results for Decision Tree Model

Parameter	Without Normalization	With Normalization
Accuracy	72.88%	73.72%
Recall	68.18%	68.64%
Precision	71%	75.35%

Thus it has been analyzed that the performance of ML techniques has been improved with normalization. But normalization impacts the performance of KNN significantly as compared to SVM and decision tree. The comparative analysis of all the undertaken techniques is presented in Fig. 5. Out of all the undertaken techniques, Decision tree performed better as it has shown the highest accuracy, precision and recall values.

Table V: Comparative Analysis of the ML Techniques

	Accuracy	Precision	Recall
KNN	71.18%	68.6%	63%
SVM	71.19%	67%	68%
Decision Tree	73.72%	75.35%	68.64%

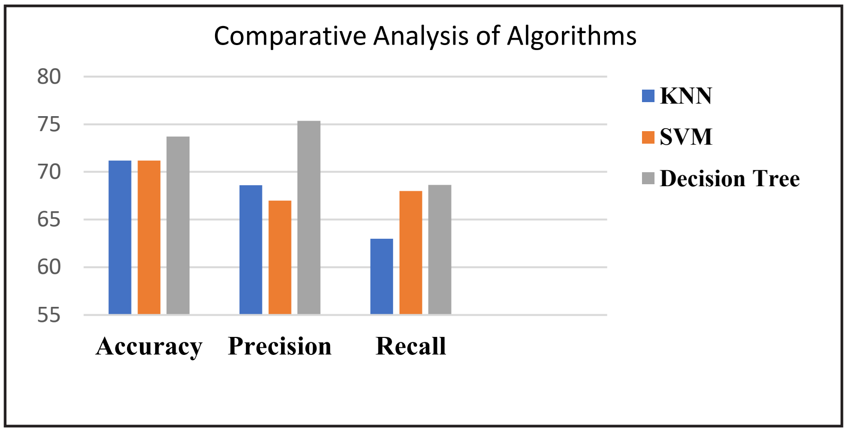

Fig. 5: Comparison of KNN, SVM and Decision Tree for Acute Appendicitis Prediction

V. Conclusion and Future Scope

The diagnosis of acute appendicitis is based primarily on a clinical examination of the signs and symptoms of the patient. Close monitoring by diagnostic imaging and laboratory testing would seem useful in situations where the diagnosis is difficult. It is a difficult task to diagnose appendicitis correctly as the signs are often indicative of certain other illnesses.

The aim of the current research was to use machine learning techniques to predict appendicitis with good accuracy and to evaluate the performance of these techniques. The experimental results show that the decision tree approach is better in predicting acute appendicitis with a diagnostic accuracy of 73.72%, as compared to KNN and SVM. In conclusion, it can be stated that machine learning techniques can predict acute appendicitis with good diagnostic accuracy.

In future work can be extended to even wider age group incorporating more features related to appendicitis. Also, urine test reports, as well as appendicitis in pregnant females, can also be included.

References

[1] M. W. Jones, R. A. Lopez, and J. G. Deppen, “Appendicitis,” *Treasure Island (FL): Stat Pearls Publishing* 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK493193/#:~:text=4%5D%5B5%5D,Epidemiology,men%20and%206.7%20%25%20for%20women

[2] T. Sarkar, “AI and machine learning for healthcare,” Towardsdatascience.com. [Online]. Available: https://towardsdatascience.com/ai-and-machine-learning-for-healthcare-7a70fb3acb67 (Accessed July 7, 2020).

[3] M. Kothainayaki, and P. Thangaraj, “Clustering and classifying diabetic data sets using k-means algorithm,” *Journal of Applied Information Science,* vol. 1, no. 1, pp. 23-27, 2013.

[4] R. Gupta, “Mining fuzzy amino acid associations in peptide sequences of Herpes simplex virus,” *Journal of Applied Information Science,* vol. 1, no. 2, pp. 10-15, 2013.

[5] S. Manonmani, R. Shanta, K. D. Kumar, S. Ishan, J. Akhilesh, and I. Joshua, “2D to 3D conversion of images using Defocus method along with Laplacian matting for improved medical diagnosis,” *Journal of Applied Information Science,* vol. 6, no. 2, pp. 6-13, 2018.

[6] H. Hashim, N. G. Mathew, K. Sabira, A. Nizamudeen, and J. Jacob, “Advanced medical diagnosis and prediction using deep learning,” *Journal of Applied Information Science,* vol. 7, no. 1, pp. 11-15, 2019.

[7] S. G. Prabhudessai, S. Gould, S. Rekhraj, P. P. Tekkis, G. Glazer, and P. Ziprin, “Artificial neural networks: Useful aid in diagnosing acute appendicitis,” *World*

Journal of Surgery, vol. 32, no. 2, pp. 305-309, 2007.

[8] E. Sivasankar, S. Rajesh, and R. Venkateswaran, "Diagnosing appendicitis using back propagation neural network and Bayesian based classifier," *International Journal of Computer Theory and Engineering*, vol. 1, no. 4, pp. 1793-8201, October 2009.

[9] H.-W. Ting, J.-T. Wu, C.-L. Chan, S.-L. Lin, and M.-H. Chen, "Decision model for acute appendicitis treatment with decision tree technology - A modification of the Alvarado scoring system," *Journal of the Chinese Medical Association,* vol. 73, no. 8, pp. 401-406, 2010.

[10] Ö. Yoldaş, M. Tez, and T. Karaca, "Artificial neural networks in the diagnosis of acute appendicitis," *The American Journal of Emergency Medicine*, vol. 30, no. 7, pp. 1245-1247, September 2011.

[11] R. Ohle, F. O'Reilly, K. K. O'Brien, T. Fahey, and B. D. Dimitrov, "The Alvarado score for predicting acute appendicitis: A systematic review," *BMC Medicine*, vol. 9, no. 139, 2011.

[12] R. Balu, and T. Devi, "Design and development of automatic appendictis detection system using sonographic image mining," *International Journal of Engineering and Innovative Technology,* vol. 1, no. 3, March 2012.

[13] R. S. Jade, U. Muddebihal M., and Naveen N., "Modified Alvarado score and its application in the diagnosis of acute appendicitis," *International Journal of Contemporary Medical Research,* vol. 3, no. 5, pp. 1398-1400, May 2016.

[14] Data Project AI Pediatric Appendicitis, OSF Data Repository, December 2019. [Online]. Available: https://osf.io/9wvys

[15] A. Uberoi, "K-Nearest Neighbor". Geeksforgeek.org. [Online]. Available: https://www.geeksforgeeks.org/k-nearest-neighbours/ (Accessed June 27, 2020).

[16] C. Liu, "A top machine learning algorithm explained: Support Vector Machines (SVM)". kdnuggets.com. [Online]. Available: https://www.kdnuggets.com/2020/03/machine-learning-algorithm-svm-explained.html (Accessed June 28, 2020).

[17] A. Sharma, "Decision Tree introduction with example". Geeksforgeek.org. [Online]. Available: https://www.geeksforgeeks.org/decision-tree-introduction-example/ (Accessed June 27, 2020).