Convolutional Neural Networks, Compared to Decision Tree, Provide Better Accuracy for Handwritten Digit Recognition

Kalakada Karthik[1], Magesh Kumar.S[2*]

[1]Research Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu, India.

[2]*Project Guide, Corresponding Author, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu, India.

ABSTRACT

Aim: The aim is to improve and develop a written digit identification to detect practically important issues in pattern recognition using Novel Convolution Neural Networks compared with Decision tree. **Materials and Methods:** Handwritten digit recognition is performed using Novel Convolution neural networks algorithm Sample size is 66 over decision tree algorithm Sample size is (N=66) with the split size using Table 1 and training using G power 80% and testing dataset 70% and 30% respectively. **Results:** The retrieval accuracy of the Novel Convolution Neural Networks classifier is (96.42%) and Decision tree is (72.35%) , There exists a statistically insignificant difference between the two groups (p=0.193; p>0.05)`. **Conclusion:** The work has confirmed that the efficiency of the Novel convolution neural network algorithm has given more accuracy value in written digit identification when compared to decision tree algorithm using artificial intelligence algorithm.

Keywords: Novel Convolution Neural Networks, Decision Tree, Recognition System, Supervised Learning, Multilayer Perceptron, Handwritten Digit.

INTRODUCTION

Handwritten digit identification is referred to as optical digit identification. The written digit can be detected through Scanners and cameras. The document's picture is divided into lines, numerals, and individual characters. OCR methods are used to recognize each digit (Hand 2020). A written digit that can scan. In a Multilayer Perceptron, the digits are converted into a machine-readable form, which can then be used for data processing such as editing or searching (Bhattacharyya et al. 2020). A written digit that can scan a document or picture file and then transform it to a machine-readable form for data processing such as editing or searching (Chakraborty et al. 2021)). In Handwritten digit identification, one has to face many difficulties in view of various styles of composing various groups of people as it's anything but an Optical digit identification by using Multilayer Perceptron (Liu, Wei, and Meng 2020). This exploration gives an extensive correlation between various Artificial Intelligence and profound clear sampling calculations with the end goal of written by hand digit identification. For this, they have utilized Scalable versions of the Support Vector Machine, Multilayer Perceptron, and Novel Convolutional Neural Network. The exactness, errors, and testing-preparation time are used to complete the correlation between these computations, which is accompanied by

plots and outlines created using for perception. The various algorithms are discussed in the literature survey. Application ofHandwritten digit identification is used in various places that can help to recognize systems that are less accurate than those found on the internet. Lexicons or spelling checkers are used to rectify the inaccuracies.

The handwritten digit identification is implemented to help customers with nearly 1448 papers published in IEEE explorer and 17,300 papers published in Google Scholar. Spiking Neural Network for Recognition of Handwritten Digits-Supervised Learning and Network Optimization (Kulkarni and Rajendran 2018). Handwritten digit recognition system on convolutional neural networks in Multilayer Perceptron (Li et al. 2020). Handwritten digit recognition based on depth neural network (Hou and Zhao 2017). Handwritten digit recognition system on an FPGA(Si and Harris 2018).(Parakh et al. 2020; Pham et al. 2021; Perumal, Antony, and Muthuramalingam 2021; Devarajan et al. 2021; Dhanraj and Rajeshkumar 2021; Uganya, Radhika, and Vijayaraj 2021; Tesfaye Jule et al. 2021; Nandhini, Ezhilarasan, and Rajeshkumar 2020; Kamath et al. 2020)

This study deals with many findings from the existing papers on the problem of classifying the strokes. In the literature survey fewer numbers of trained dataset approaches are used which leads to a decrease in accuracy. The aim of this paper is to train with more data which results in increased accuracy with the classifier of mean prediction in the written digit identification, which is classified into two groups as Novel Convolution Neural Networks in Supervised Learning, compared with the order of other algorithms using Decision tree algorithm.

MATERIALS AND METHODS

The setup with the research has been performed in the Data Analytical Laboratory of the Department of Computer Science and Engineering in Saveetha School of Engineering ,Saveetha Institute of Medical And Technical Sciences. Which hasn't got any ethical approval yet. The project mainly depends on two algorithms one is for the base and another for comparison in Supervised Learning, which is classified into two groups like Novel Convolution Neural Networks, and Decision tree with two samples size of 66 as shown in Table 1 which is a total of 132 rows using G power 80 % ,which is done using pretest power of 0.001.

Convolutional Neural Network

Convolutional Neural Network is a system of feed-forward neural architecture used in artificial intelligence. It's a standard tool for image recognition. It utilizes a multidimensional array to represent input data. It works well although there is a lot of labeled data. It extracts a perceptron from each and every part of the input images.It is the layer where it contributes to our model. The absolute number of provisions in our information size is equivalent in relation to the number of neurons in the layer.Depending on our model and data amount, the hidden layer will appear. Each layer has a different number of neurons that exceeds the number of characteristics. Each layer's output is calculated by multiplying the learnable weights' output for the previous layer of the layers in a matrix.The hidden layer's output is then passed into a logistic function like sigmoid

or softmax, which turns it into probability scores for each class.

The pooling layer works similarly to the traditional layer because it reduces the dimension of the feature of the convolution feature. The amount of processing capacity required to process the data is lowered by lowering the dimensionality of the data. It also aids the model's training phase by identifying rotating and spatial invariant dominant features.

Pseudocode:

```
For x from 1 to q do          —inter output
For y from 1 to r do          --intra output
For w from 1 to P do
For a from 1 to R do
temp =0;
For xx from 1 to t do
For yy from 1 to t do
   Temp = temp + D[xx][yy]* A[y][s*(r - 1) +xx]*(c-1)+b
End for
B[x][r][c]= f(B[x][r][c] +tmp
If y == m
    B[x][r][c]= f(B[x][r][c] +bias
End if
End for
End for
End for
End for
```

Decision Tree

The Decision Tree algorithm is part of the supervised learning algorithm family. The goal of a choice tree is to take in essential choice tree rules from past information and use them to develop a planning model that can predict the target variable's class or value. A decision tree has two types of variables categorical and continuous.

The way essential elements are made has a big impact on how precise a tree is. For characterization and relapse trees, it has its own set of rules. It uses a number of formulas to divide hubs into at least two sub-hubs. It may create the virtue of a hub in terms of the target variable and then select the split that results in the most homogenous sub-hubs. For solving the problem, Given training vectors xi ∈ Rn, i=1,…, l and a label vector y ∈ Rl, in order to solve a problem, Given xiRn, i=1,..., l training vectors and a label vector yRl, allow Qm to compare the data at node m with the data from Nm samples as shown in equation 1.

Qmleft() and Qmright() subset Qmleft() = Qmleft(θ) and Qmright(θ) subset Qmleft(θ) = {(x,y) |xj<=tm } Qmright(θ) = Qm \ Qmlef t(θ) ---------- (1)

The quality of a candidate split of node m is then computed using an impurity function or loss function H(), the choice of which depends on the task being solved (classification, regression)G(Qm,θ)=NmleftNmH(Qmleft(θ))+NmrightNmH(Qmright(θ)). Choose the parameters that will result in the least amount of contamination. θ∗=argminθf_0G(Qm,θ). Recurse until the maximum permissible depth is achieved for subsets Qmleft() and Qmright(), Nm m is a terminal node, and the predicted probe for this region is set to pmk. The following are some examples of impurity measurements.

Gini:H(Qm)=∑kpmk(1−pmk) -------- (2)

Entropy:H(Qm)=−∑kpmg log[f0](pmk) ------ (3)

Pseudocode:

```
DecTree(Sample A, Features B)
if stopping_conditions(A,B) = true then{
    Leaf = create Node()
    LeafLabel =classify(a)
    Return leaf
}
{
Root = createNode()
Root.test_conition = findBestSpilt(A,B)
Y ={y\y a possible output foot.test_condition}{
    C ={ c\root.test_condition(y) = y and c$ C};
    Child = TreeGrowth(Y,A);
    Add Child;
}
Return root ;
```

The model is tested in the setup of hardware requirements as I 5 processor, 8GB RAM, and 512 SSD ROM is used. The software configuration in windows 10 and Jupiter or google collab and pre-installed chrome and with the help of MSEXCAL. The code ran on a Jupyter notebook and the process of testing included downloading the required dataset to run the code to get an accurate result.

Statistical Analysis

The statistical software which is used for doing analysis in IBM SPSS version 22 with 64 bit, which is an analysis software which is done by uploading dataset to the software which gives the output as independent variable test code, second the written id, third passage id, and fourth the sentence id and dependent variables images pixel(0 to 9) with the accuracy areas the output for the given models CNN and Decision Tree (Bright 2020).

RESULTS

For the results, it has been observed that the accuracy of the NovelConvolution Neural Networks algorithm is approximately 96.42% and the Decision tree algorithm is approximately 72.35%. The accuracy of the algorithms differs in decimals when it executes them with different sizes as noted in Table 1. Both the algorithms were executed with different test sizes as noted sample size (N=66). The performance of the algorithms is compared by using Group Statistics such as accuracy and Nord in Table 1. From Group Statistics, the mean accuracy and standard deviation for the Novel Convolution Neural Networks algorithm 90.52% and Decision tree algorithm are (71.86%). By performing statistical analysis of 66 samples, Novel Convolution Neural Networks algorithm got mean value 90.52 and its standard deviation with (1.00590) standard error (0.12382) while Decision tree algorithm obtained mean value (71.86) and its standard deviation with (0.92320) standard error (0.11364) in Table 2. There exists a statistically insignificant difference between the two groups (p=0.193; p>0.05). An Independent sample test was done and compared the algorithms and the results standard error difference value(0.16806). A bar graph was plotted by the results in Fig. 1. The experimental results show that the CNN algorithm is significantly better than the Decision Tree algorithm in Multilayer Perceptron.

DISCUSSION

The analysis of the algorithm has been done with Table 2 representing the group statistics and Table 3 representing the independent variables and bar-graph which represents the compassion of two algorithms with the accuracy percentages of (90.45%) and (71.86%) for CNN and Decision tree respectively in Figure 1.There exists a statistically insignificant difference between the two groups ($p=0.193$; $p>0.05$).

There are more research, which is related to the similar study of proposed research where the findings are ”Improved handwritten digit recognition using conventional neural network” (Ahlawat et al. 2020), “Hybrid CNN-SVM classifier for handwritten digit recognition” [A Choudhary - Procedia Computer Science, 2020-Elsevier] (Ahlawat and Choudhary 2020). “Handwritten digit recognition with artificial neural network” [KH Tohidual Islam, Ghulam Mujtaba, Ram Gopal Raj, 2017 - (ICE2T) - IEEE] (Islam et al. 2017), “Design and Implementation of handwritten digit recognition system based on template method using Scalable method” Yang Zhiqi, Fu Kai, 2018 IEEE 3rd(IAEAC)] (Zhiqi and Kai 2018).

Some opposing findings were also there to find the handwritten digits using Artificial Intelligence is that “An adaptive deep Q-learning strategy for handwritten digit recognition” [J Qiao, G wang, w Li, M Chen - Neural network,2018- Elsevier] (Qiao et al. 2018). The limitation that is faced while this project is handwritten digit recognition is a big challenge with a limited number of attributes. To improve the image attributes feature enhancement by using handwritten Algorithm gets good accuracy.

CONCLUSION

In the proposed work, the accuracy percentage of predicting for Novel Convolution Neural Networks 96.42% is better than when compared with Decision tree 72.35%. The work has confirmed that the efficiency of the convolution neural network algorithm has given more accuracy value in written digit identification when compared to decision tree algorithm using artificial intelligence algorithm. CNN algorithm improved the accuracy value.

DECLARATIONS

Conflict of interest

No conflict of interest in this manuscript.

Authors Contribution

Author KK was involved in dataset collection, algorithm development, image analysis, and manuscript writing. Author TV was involved in the validation and review of the manuscript.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science (Formally known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Findings

We would like to thank the following organization for providing financial support that enabled us to complete this study.

1. INFI DATA Pvt. Ltd.
2. Saveetha university.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCE

1. Ahlawat, Savita, and Amit Choudhary. 2020. "Hybrid CNN-SVM Classifier for Handwritten Digit Recognition." *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2020.03.309.
2. Ahlawat, Savita, Amit Choudhary, Anand Nayyar, Saurabh Singh, and Byungun Yoon. 2020. "Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN)." *Sensors*. https://doi.org/10.3390/s20123344.
3. Bright, Steven. 2020. *SPSS: The Ultimate Data Analysis Tool*. Lulu Press, Inc.
4. Chakraborty, Anuran, Rajonya De, Samir Malakar, Friedhelm Schwenker, and Ram Sarkar. 2021. "Handwritten Digit String Recognition Using Deep Autoencoder Based Segmentation and ResNet Based Recognition Approach." *2020 25th International Conference on Pattern Recognition (ICPR)*. https://doi.org/10.1109/icpr48806.2021.9412198.
5. Devarajan, Yuvarajan, Beemkumar Nagappan, Gautam Choubey, Suresh Vellaiyan, and Kulmani Mehar. 2021. "Renewable Pathway and Twin Fueling Approach on Ignition Analysis of a Dual-Fuelled Compression Ignition Engine." *Energy & Fuels: An American Chemical Society Journal* 35 (12): 9930–36.
6. Dhanraj, Ganapathy, and Shanmugam Rajeshkumar. 2021. "Anticariogenic Effect of Selenium Nanoparticles Synthesized Using Brassica Oleracea." *Journal of Nanomaterials* 2021 (July). https://doi.org/10.1155/2021/8115585.
7. Hand, David J. 2020. "Artificial Intelligence Frontiers in Statistics." https://doi.org/10.1201/9781003059875.
8. Hou, Yawei, and Huailin Zhao. 2017. "Handwritten Digit Recognition Based on Depth Neural Network." *2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*. https://doi.org/10.1109/iciibms.2017.8279710.
9. Islam, Kh Tohidul, Ghulam Mujtaba, Ram Gopal Raj, and Henry Friday Nweke. 2017. "Handwritten Digits Recognition with Artificial Neural Network." *2017 International Conference on Engineering Technology and Technopreneurship (ICE2T)*. https://doi.org/10.1109/ice2t.2017.8215993.
10. Kamath, S. Manjunath, K. Sridhar, D. Jaison, V. Gopinath, B. K. Mohamed Ibrahim, Nilkantha Gupta, A. Sundaram, P. Sivaperumal, S. Padmapriya, and S. Shantanu Patil. 2020. "Fabrication of Tri-Layered Electrospun Polycaprolactone Mats with Improved Sustained Drug Release Profile." *Scientific Reports* 10 (1): 18179.
11. Kulkarni, Shruti R., and Bipin Rajendran. 2018. "Spiking Neural Networks for Handwritten Digit recognition—Supervised Learning and Network Optimization." *Neural Networks*. https://doi.org/10.1016/j.neunet.2018.03.019.
12. Li, Jinze, Gongbo Sun, Leiye Yi, Qian Cao, Fusen Liang, and Yu Sun. 2020. "Handwritten Digit Recognition System Based on Convolutional Neural Network." *2020 IEEE International Conference on Advances*

in Electrical Engineering and Computer Applications(AEECA). https://doi.org/10.1109/aeeca49918.2020.9213619.

13. Liu, Wenfei, Jingcheng Wei, and Qingmin Meng. 2020. "Comparisions on KNN, SVM, BP and the CNN for Handwritten Digit Recognition." *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications(AEECA)*. https://doi.org/10.1109/aeeca49918.2020.9213482.
14. Nandhini, Joseph T., Devaraj Ezhilarasan, and Shanmugam Rajeshkumar. 2020. "An Ecofriendly Synthesized Gold Nanoparticles Induces Cytotoxicity via Apoptosis in HepG2 Cells." *Environmental Toxicology*, August. https://doi.org/10.1002/tox.23007.
15. Parakh, Mayank K., ShriraamUlaganambi, Nisha Ashifa, Reshma Premkumar, and Amit L. Jain. 2020. "Oral Potentially Malignant Disorders: Clinical Diagnosis and Current Screening Aids: A Narrative Review." *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation* 29 (1): 65–72.
16. Perumal, Karthikeyan, Joseph Antony, and Subagunasekar Muthuramalingam. 2021. "Heavy Metal Pollutants and Their Spatial Distribution in Surface Sediments from Thondi Coast, Palk Bay, South India." *Environmental Sciences Europe* 33 (1). https://doi.org/10.1186/s12302-021-00501-2.
17. Pham, Quoc Hoa, SupatChupradit, Gunawan Widjaja, Muataz S. Alhassan, Rustem Magizov, Yasser Fakri Mustafa, Aravindhan Surendar, AmirzhanKassenov, Zeinab Arzehgar, and WanichSuksatan. 2021. "The Effects of Ni or Nb Additions on the Relaxation Behavior of Zr55Cu35Al10 Metallic Glass." *Materials Today Communications* 29 (December): 102909.
18. Qiao, Junfei, Gongming Wang, Wenjing Li, and Min Chen. 2018. "An Adaptive Deep Q-Learning Strategy for Handwritten Digit Recognition." *Neural Networks: The Official Journal of the International Neural Network Society* 107 (November): 61–71.
19. Si, Jiong, and Sarah L. Harris. 2018. "Handwritten Digit Recognition System on an FPGA." *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. https://doi.org/10.1109/ccwc.2018.8301757.
20. Tesfaye Jule, Leta, Krishnaraj Ramaswamy, Nagaraj Nagaprasad, Vigneshwaran Shanmugam, and Venkataraman Vignesh. 2021. "Design and Analysis of Serial Drilled Hole in Composite Material." *Materials Today: Proceedings* 45 (January): 5759–63.
21. Uganya, G., Radhika, and N. Vijayaraj. 2021. "A Survey on Internet of Things: Applications, Recent Issues, Attacks, and Security Mechanisms." *Journal of Circuits Systems and Computers* 30 (05): 2130006.
22. Zhiqi, Yang, and Fu Kai. 2018. "Design and Implementation of Handwritten Digit Recognition System Based on Template Method." *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*.

https://doi.org/10.1109/iaeac.2018.8577760.

23. Ahlawat, Savita, and Amit Choudhary. 2020. "Hybrid CNN-SVM Classifier for Handwritten Digit Recognition." Procedia Computer Science. https://doi.org/10.1016/j.procs.2020.03.309.
24. Ahlawat, Savita, Amit Choudhary, Anand Nayyar, Saurabh Singh, and Byungun Yoon. 2020. "Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN)." Sensors. https://doi.org/10.3390/s20123344.
25. Bhattacharyya, Siddhartha, Vaclav Snasel, Aboul Ella Hassanien, SatadalSaha, and B. K. Tripathy. 2020. Deep Learning: Research and Applications. Walter de Gruyter GmbH & Co KG.
26. Chakraborty, Anuran, Rajonya De, Samir Malakar, Friedhelm Schwenker, and Ram Sarkar. 2021. "Handwritten Digit String Recognition Using Deep Autoencoder Based Segmentation and ResNet Based Recognition Approach." 2020 25th International Conference on Pattern Recognition (ICPR). https://doi.org/10.1109/icpr48806.2021.9412198.
27. Hand, David J. 2020. "Artificial Intelligence Frontiers in Statistics." https://doi.org/10.1201/9781003059875.
28. Hou, Yawei, and Huailin Zhao. 2017. "Handwritten Digit Recognition Based on Depth Neural Network." 2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). https://doi.org/10.1109/iciibms.2017.8279710.
29. Islam, KhTohidul, Ghulam Mujtaba, Ram Gopal Raj, and Henry Friday Nweke. 2017. "Handwritten Digits Recognition with Artificial Neural Network." 2017 International Conference on Engineering Technology and Technopreneurship (ICE2T). https://doi.org/10.1109/ice2t.2017.8215993.
30. Kulkarni, Shruti R., and Bipin Rajendran. 2018. "Spiking Neural Networks for Handwritten Digit recognition—Supervised Learning and Network Optimization." Neural Networks. https://doi.org/10.1016/j.neunet.2018.03.019.
31. Li, Jinze, Gongbo Sun, Leiye Yi, Qian Cao, Fusen Liang, and Yu Sun. 2020. "Handwritten Digit Recognition System Based on Convolutional Neural Network." 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications(AEECA). https://doi.org/10.1109/aeeca49918.2020.9213619.

TABLES AND FIGURES

Table 1: Comparing accuracy values with the different sample sizes. It Represents the diabetes prediction, the accuracy of the Convolutional Neural Networks Algorithm (96.42%), and the Decision Tree algorithm (72.35%).

S.No.	Sample Size	Convolutional Neural Networks Accuracy(%)	Decision Tree Accuracy(%)
1	768	96.42	72.35
2	745	96.23	72.34
3	715	96.00	72.26
4	690	95.54	72.13
5	677	95.23	72.10
6	656	95.00	72.09
7	639	95.68	72.32
8	624	95.89	72.16
9	615	94.26	72.06
10	600	94.23	72.01
11	586	94.16	72.00
12	571	94.67	71.95
13	555	94.00	71.86
14	542	93.25	71.72
15	530	93.75	71.67
16	538	93.24	71.80
17	517	73.16	71.23
18	516	73.08	71.19
19	509	73.64	71.09
20	500	73.04	71

Table 2: The CNN has 90.52% method and grouped statistics were compared using group statistics for recorded data from simulation for 66 iterations 71.86%. In comparison, the Decision tree algorithm has a high level of accuracy.

	Algorithm	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	CNN	20	90.52	1.00590	.12382
	DT	20	71.86	.92320	.11364

Table 3: The logged data from the simulation, an independent sample test was performed for 66 iterations to set the confidence interval to 95% and the threshold of significance to 0.05. There exists a statistically insignificant difference between the two groups (p=0.193; p>0.05).

		Levene's Test for Equality of Variance		**T-test for Equality of Means**						
		f	**Sig**	**t**	**df**	**Sig.(2-tailed)**	**Mean Difference**	**Std.Error Difference**	**95% Confidence of the Differences**	
									Lower	**Upper**
ACCURACY	Equal variances assumed	1.710	.193	143.230	130	<.001	24.07136	.16806	23.73888	24.40385
	Equal variances not assumed			143.230	129.055	<.001	24.07136	.16806	23.73885	24.40387

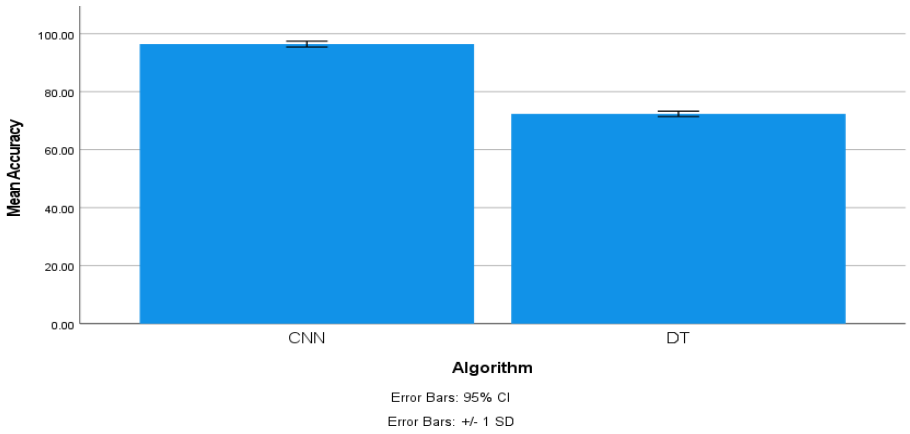

Fig. 1: Performance comparison between CNN(96.42%) and Decision tree(72.35%). The mean precision of CNN and the standard deviation of DT is better than CNN. X-axis CNN Vs DT algorithm Y-axis Mean precision. Error Bar +/-1 SD, Error Bar 95%.