2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 978-1-6654-3789-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICACITE53722.2022.9823806

MNIST Handwritten Digit Recognition using Machine Learning

Elizabeth Rani. G
Assistant Professor
Department of CSE
Erode Sengunthar Engineering College
Erode, Tamilnadu, India.
elizarn292@gmail.com

Sakthimohan. M
Associate Professor
Department of ECE
Erode Sengunthar Engineering College
Erode, Tamilnadu, India.
sakthimohan.phd@gmail.com

Abhigna Reddy. G
Department of CSE
Kalasalingam Academy of Research and Education
Anand Nagar, Krishnankoil, India
9920004158@klu.ac.in

Selvalakshmi. D
Department of CSE
Kalasalingam Academy of Research and Education
Anand Nagar, Krishnankoil, India
9920004122@klu.ac.in

Thomalika keerthi
Department of CSE
Kalasalingam Academy of Research and Education
Anand Nagar, Krishnankoil, India
9920004136@klu.ac.in

Raja Sekar. R
Department of CSE
Kalasalingam Academy of Research and Education
Anand Nagar, Krishnankoil, India.
r.rajasekar@klu.ac.in

Abstract - Identification of the MNIST database that will be in handwritten digit can be recognized by the machine. It can be recognized that the human handwritten form into the machine language. Here we are using some machine learning algorithm that is Convolutional Neural Network (CNN). MNIST database consist of nine digits that is 0 to 9. It can be used to change over manually written digits into machine lucid arrangements. The primary target of this work is to guarantee successful and dependable methodologies for acknowledgment of written by hand digits and make banking tasks more straightforward and mistake free. The MNIST database will be identified by the machine very accurately. Here we are finding the handwritten digit accuracy.

Keywords — MNIST handwritten digit recognition, Image recognition, digit recognition, Convolutional neural network (CNN).

I. Introduction

The MNIST information base is an acronym that represents the Modified National Institute of Standard and Technology dataset. MNIST manually written digit acknowledgment is a dataset of 60,000 little square 28 * 28-pixel grayscale picture of transcribed single digits somewhere in the range of 0 and 9. The assignment is to order the given picture of a manually written digit into one of 10 classes addressing whole number qualities from 0 to 9 comprehensively. The transcribed digit are not dependably of a similar size, width, direction and legitimized to edges as they contrast from composing of individual to individual, so broad issue would be while arranging the digits like 1 and 7 , 5 and 6 , 3 and 8 , 2 and 5 , 2 and 7 and so forth… .. The picture of written by hand digits as 10 digits (0 to 9). The MNIST Handwritten Digits is considered as the "Hi World" of Computer Vision. The greater part of the standard executions of neural organizations accomplishes a precision of ~ (98 - 99) percent in accurately characterizing the transcribed digits. Past this number, each and every decimal expansion in the precision rate is hard. The Handwritten digit acknowledgment is the answer for the issue which utilizes the picture of a digit and perceives the digit present in the picture. It is a hard errand for the machine in light of the fact that transcribed digits are somewhat flawed and can be made with many flavors. Here we are utilizing the Convolutional Neural Network (CNN) calculation, is a sort of Artificial Neural Network utilized in picture acknowledgment and handling that is extraordinarily intended to deal with pixel information. A neural organization is an arrangement of equipment and/or programming designed after the activity of neurons in the human cerebrum.

Perceiving written by hand digits is critical, and it has an assortment of uses from web-based penmanship acknowledgment on 21st-century cell phones and tablets. To deal with postal zip numbers extricated from mail or letterheads, Numeric sections in administrative work, (for example, government forms) and bank check sums structures) finished up manually, or to distinguish tags consequently [1], etc. While dealing with a task, there are a few troubles that should be survives, to observe an answer for this issue, the digits are composed manually. changes with strokes, size, thickness, and direction and furthermore the separation from the edges at last increments. So to perceive those are a lot of complex. To do as such, [2] few endeavors at the transcribed digits have been perceived utilizing by consolidating SVM with rule-based thinking, ANN [3] was made. Profound Neural Networks with numerous segments [4] for Image Classification were utilized.

Fig. 1. Sample image from MNIST dataset

Fig 1 shows is the model picture of the MNIST dataset. It comprises of 10 numbers that is 0 to 9. From this we can observe the drawn picture of the number is recognized plainly and precisely.

The objective was to construct a model that could Categories a digit as per its example by utilizing CNN to perceive a manually written digit with a comparable example. In this examination, we propose an original CNN model for written by hand digit recognizable proof from arbitrary pictures utilizing the MNIST dataset that accomplishes the best outcomes. The pictures are of English manually written numerals that were assembled from an assortment of sources and downsized to 28x28 to squeeze into the model. To separate the highlights of the digits in the picture, we fostered our model with four convolutional layers [5], with two max-pooling layers after each two convolutional layers. [6]

II. LITERATURE SURVEY

Anuj Dutt in his paper showed that using Deep learning frameworks, he had the abilities to get a very high accurac . By using the Convolutional Neural Network (CNN) with keras and theano as backend, he was getting the exactness of 98.72%. What's more, execution of CNN using TensorFlow gives an incredibly better result of 99.70%. Despite the fact that the complexity of the system and codes seems, by all accounts, to be more when diverged from normal Machine learning calculations yet the exactness he got is progressively self-evident.

In a paper distributed by Saeed AL - Mansoori, Multi Layer Perceptron (MLP) Neural Network was executed to perceive and foresee transcribed digits from 0 to 9.

Profound learning is a kind of AI technique that works with an assortment of information portrayals. [7-9] Algorithms are made by concentrating on the construction and tasks of the human mind. Directed, semi-administered, and solo learning are a wide range of profound learning. Profound neural organizations, Deep conviction organizations, and Recurrent neural organizations are instances of profound learning structures utilized in regular language handling, sound handling PC vision, discourse acknowledgment, informal community separating, machine interpretation, bioinformatics, and clinical picture handling. [10-12]. They have accomplished outcomes that are identical to, and now and again far and away superior to, people. [13-15].

III. METHODOLOGY

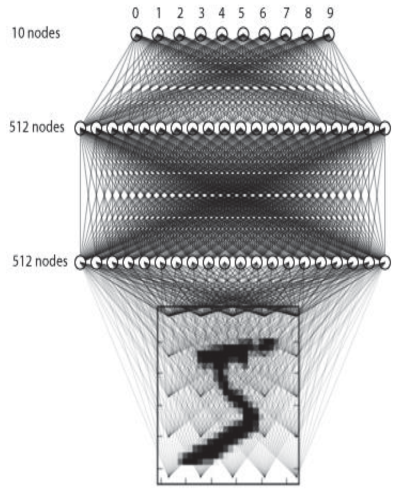

Fig. 2. Convolutional Neural Network Architecture

The details parameters of CNN are given by

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_4 (Conv2D)            (None, 28, 28, 64)        640
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 14, 14, 64)        0
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 14, 14, 64)        36928
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 7, 7, 64)          0
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 7, 7, 64)          36928
_________________________________________________________________
flatten_2 (Flatten)          (None, 3136)              0
_________________________________________________________________
dropout_2 (Dropout)          (None, 3136)              0
_________________________________________________________________
dense_2 (Dense)              (None, 10)                31370
_________________________________________________________________
activation_2 (Activation)    (None, 10)                0
=================================================================
Total params: 105,866
Trainable params: 105,866
Non-trainable params: 0
```

Fig. 3. Parameters of CNN

A. *Input Layer*

We construct our network to take a 784-length vector rather than a 28 x 28 matrix in table 1. After that, each image must be moulded (or flattened) into a vector.

We'll also change the range of the inputs to [0-1] rather than [0-255]. It's often a good idea to normalise inputs so that any additional dimensions (for alternative network designs) are on the same scale [16] which is shown in fig 2 and 3.

B. *Convolutional Layer 1*

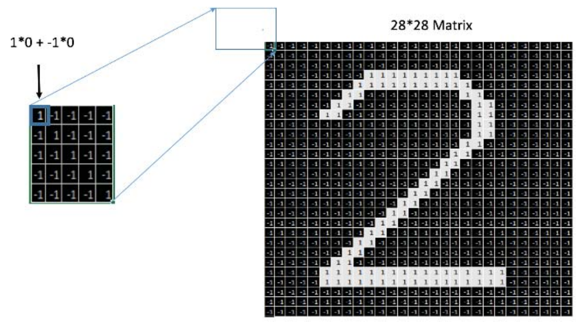

Fig. 4. Convolve with weight tensor and add biases

Kernel weight and bias are defined.

Here, we define a kernel. The filter/kernel is 5x5 in size; the input channels are 1 (grayscale); and we require 32 separate feature maps (32 feature maps implies 32 different filters are applied to each image) in fig 4.

As a result, the convolution layer's output would be 28x28x32) [17-20]. We'll make a filter / kernel tensor with the shape code [filter width, filter height, out channels, in channels]

Input to Convolutional Layer:

We use conv2d to construct a convolutional layer. Given fourdimensional input and filter tensors, it computes a two-dimensional convolution.

Shape tensor [in height, batch, in channels, in width] X [batch size,28,28, 1]

[in channels, out channels filter height, filter width] is a filter / kernel tensor. W has a size of [1, 1, 1, 1] and stride is [5, 5, 1, 32].

The "kernel window" is shifted over the input tensor by the convolutional layer. The convolution acts on a 2D window on the height and width dimensions because the input tensor has four dimensions: [batch, height, width, channels]. strides specify the amount by which the window moves in each dimension. We set the stride to 1 because the first and last dimensions [21, 22].

C. Process

Filter 2-D matrix is converted with the shape [5*5*1,32]. - Creates a virtual tensor with the shape '[batch, 28, 28, 5*5*1]' by ex-tracting image patches from the input tensor.

- Right-multiplies the filter matrix and the imagery vector for each batch.

D. Output

Figure 5 shows A 'Tensor' (a 2-D convolution) of size tf as an output shape(?, 28, 28, 32) tensor 'add 7:0'

The first convolution layer produces 32 [28x28] pictures as output. The resulting im-age's volume/depth is 32 in this case.

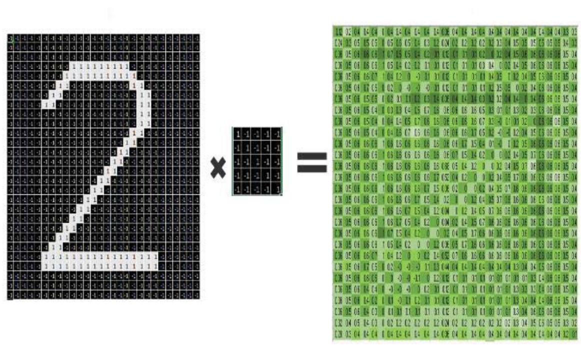

Fig. 5. Output of Convolutional Layer 1

Use the Re-LU Activation Function to get started.

In this stage, simply run through all of the outputs of the convolution layer, convolve1, and replace any negative numbers with 0. It's known as the ReLU Activation Function.

Let f(x)=max be a ReLU activation function (0,x).

E. Using maximum pooling

Non-linear down-sampling is a type of max pooling. It divides the input image into rec-tangles and then calculates the greatest value for each region.

To do maximum pooling, we'll utilize the tf.nn.max pool function. Size of a kernal: 2x2 (if the window is a 2x2 matrix, it would result in one output pixel) which is shown in fig 6

Strides: determines the kernel's sliding behav-iour. In this example, it will advance two pix-els every time, ensuring that it does not over-lap. The input would be a 28x28x32 matrix, and the output would be a 14x14x32 matrix [23, 24].

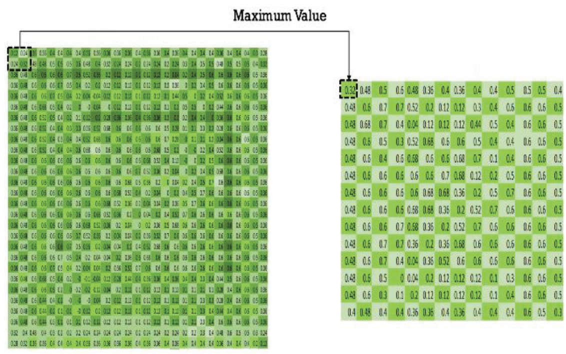

Fig. 6. Output After Applying Max Pooling

F. Convolutional Layer 2

Kernel Weights and Biases in Convolutional Layer 2

In this layer, we use convolution once more. Let's take a look at the kernel for the second layer:

5x5 filter/kernel (25 pixels)

32 input channels (from the 1st Conv layer, we had 32 feature maps)

There are 64 output feature maps.

Similarly to Layer 1, in Layer 2, we apply ReLu function and Max Pooling operation.The output of Layer 2 is 64 matrix of [7x7].

G. Fully Connected Layer

To use the Softmax and build the probabilities at the end, you'll need a completely connected layer. The high-level filtered images from the preceding layer, i.e. all 64 matrices, are converted to a flat array by fully connected layers.

So, each [7x7] matrix will be converted to a [49x1] matrix, and then all 64 matrices will be joined to form a [3136x1] array. We'll join it to another layer with the dimensions [1024x1]. As a result, the weight difference between these two layers will be [3136x1024] in fig 7.

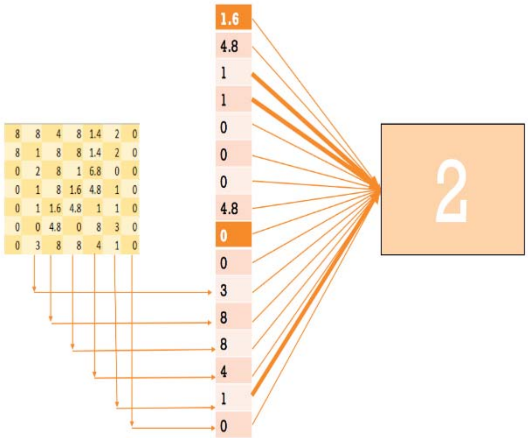

Fig. 7. Output from Fully Connected Layer

IV. RESULTS AND DISCUSSIONS

From this we can find the drawn image of MNIST dataset digit can be identified by the machine readable format . This is very hard to machine to convert it into the machine language so we are using the Machine Learning Algorithm i.e. Convolutional Neural Network(CNN) . This algorithm is used to recognize image of handwritten digit image into machine readable format. Some of the example outputs are given below in fig 8.

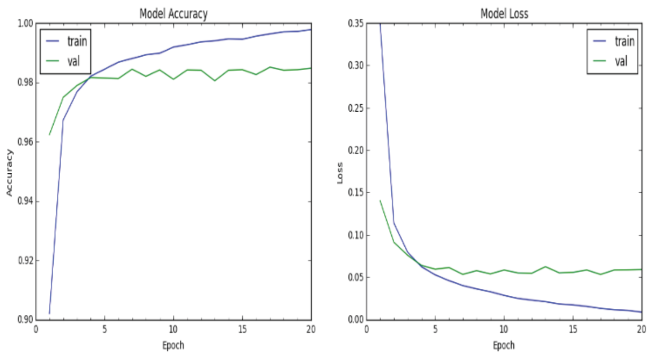

Fig. 8. Accuracy and Error Rate using Simple CNN

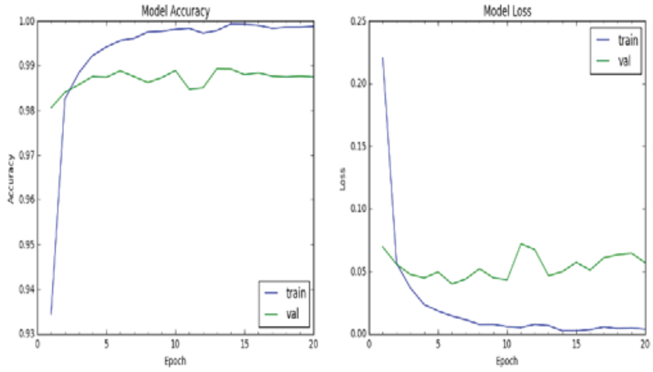

Fig. 9. Accuracy and Error Rate using Deep CNN

The above image represents the accuracy of the digit drawn by the user. The user had drawn that the digit is 2 and it was identified by the computer and it also gives the accuracy of the given digit is 98.68% in Simple CNN and 99.23% using Deep CNN. Also, the corresponding error rate were shown in above Figure 9.

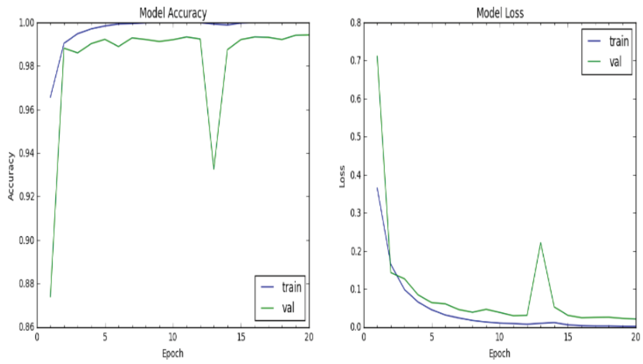

Fig. 10. Accuracy and Error Rate using Deep CNN after Batch Normalization

The above image Fig 10 represents the accuracy of the digit drawn by the user using batch normalization. The user had drawn that the digit is 2 and it was identified by the computer, and it also gives the accuracy of the given digit is 99.30%. The time taken by simple CNN, Deep CNN and Batch Normalized CNN were 142.17 second, 156.23 seconds, 179.12 seconds. It is better that accuracy obtained in [10-13]

V. CONCLUSIONS

The initial stage in enormous area of Artificial Intelligence (AI) and Computer Vision is handwritten digit recognition. CNN outperforms alternative classifiers, as evidenced by the results of the trial. With more convolution layers and buried neurons, the findings may be made more precise. This is a common dataset for evaluating classifier performance. There are three phases to Handwritten Digit Recognition (HDR). The first step is pre-processing, which involves converting the dataset to binary format and applying image processing on it. The picture is divided into several parts in the second phase, segmentation. Third step is feature extraction, which involves identifying picture characteristics. The use of CNN greatly improves the HDR results.

REFERENCES

[1] Z Selmi, M Ben Halima, A M. Alimi, "Deep Learning System for Automatic License Plate Detection and Recognition", 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017, pp. 1132-1138, doi.org/10.1109/ICDAR.2017.187

[2] Y Bengio, A Courville, P Vincent, "Representation Learning: A Review and New Perspectives", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 35, issue 8, 2018, pp 1798–1828, doi.org/10.11,9/tpami.2013.50.

[3] J Schmidhuber, "Deep Learning in Neural Networks: An Overview", Neural Networks.Vol 61, 2015, pp 85–117, doi.org/10.1016/j.neunet.2014.09.003

[4] Y Bengio, Y LeCun, G Hinton, "Deep Learning". Nature 521, 2015, pp. 436–444, doi.org/10.1038/nature14539

[5] P Le Callet, C Viard-Gaudin, D Barba, "A Convolutional Neural Network Approach for Objective Video Quality Assessment", IEEE Transactions on Neural Networks, 2006, Vol 17 Issue 5, pp. 1316–1327, doi.org/10.1109/TNN.2006.879766.

[6] Simard, Patrice, D Steinkraus, and J C. Platt. "Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis." In ICDAR, vol. 3, pp. 958–962. 2003.

[7] LeCun, Yann, Bengio, Yoshua, "Convolutional networks for images, speech, and time series". In Arbib, 4 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN) 605 Michael A. The handbook of brain theory and neural networks, edition 2, 1995, pp. 276–278.

[8] G. E. Rani, R. Murugeswari and M. Sakthimohan, "The innovative secrecy measure for data broadcasting," 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2017, pp. 1-6, doi: 10.1109/ITCOSP.2017.8303113.

[9] G. E. Rani, G. S. Reddy, P. Obulareddy, S. S. R. Harshith and M. Sakthimohan, "Web Application For Community Question Answering," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021, pp. 670-673, doi: 10.1109/ICACITE51222.2021.9404605..

[10] G. Elizabeth Rani, G. Narasimha Murthy, M. Abhiram, H. Mohan, T. Singh Naik and M. Sakthimohan, "An Automated Airlines Reservation Prediction System Using BlockChain Technology," 2021 Sixth International Conference on Image Information Processing (ICIIP), 2021, pp. 224-228, doi: 10.1109/ICIIP53038.2021.9702587

[11] G. E. Rani, A. T. V. Reddy, V. K. Vardhan, A. S. S. Harsha and M. Sakthimohan, "Machine Learning based Cibil Verification System," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 780-782, doi: 10.1109/ICSSIT48917.2020.9214195.

[12] G. Dhanabalan, V. Karutharaja and M. Sakthimohan, "Realization of Resource Efficient Block RAM Based Eight Bit Adder in FPGA," 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2019, pp. 1-5, doi: 10.1109/INCOS45849.2019.8951332.

[13] M. Sakthimohan., P. Ruchitha, K. J. Harshitha, B. Tharuni, G. Elizabeth Rani. and J. Deny., "Trigonous Sheilding System For Women," 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), 2021, pp. 397-401, doi: 10.1109/ISPCC53510.2021.9609360.

[14] S. M, E. R. G, B. G. Reddy, S. L. Reddy and V. Chennareddy, "Wireless Power Transmission Science Model," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), 2021, pp. 1-5, doi: 10.1109/ICESC51422.2021.9532606.

[15] M. Sakthimohan, M. K. Pasupathi, V. Gopal, N. Gosia and R. G. Elizabeth, "IOT Established Agile Parking System," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 2021, pp. 637-641, Doi: 10.1109/ICSPC51351.2021.9451732

[16] G. E. Rani, S. Deetshana, K. Y. Naidu, M. Sakthimohan and T. Sarmili, "Automated Interactive Irrigation System - IoT Based Approach," 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2019, pp. 1-4, doi: 10.1109/INCOS45849.2019.8951382.

[17] Sakthimohan. M, R. Prasanth, R. G, H. Narayanan, and E. R. G, "Optimized Transmission Line Flaw Disclosure and Inkling System Based on IOT," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), 2021, pp. 642-646, Doi: 10.1109/ICSPC51351.2021.9451715

[18] Inundate, Madhavi Kannan, Alisha RM, Vijayalakshmi P, Shrestha sin, "Plan and Development of an IOT based mostly wearable gadget for the Safety and Security of ladies and young lady kids", IEEE, 2016.

[19] M. Sakthimohan, J. Deny, G. Elizabeth Rani et al., IOT based shrewd agronomy method, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.11.096

[20] Sakthimohan M, Deny J, An Optimistic Design of 16-Tap FIR Filter with Radix-4 Booth Multiplier Using Improved Booth Recoding Algorithm, Microprocessors and Microsystems, https://doi.org/10.1016/j.micpro.2020.103453

[21] Sakthimohan M., Deny J. (2021) An Efficient Design of 8 * 8 Wallace Tree Multiplier Using 2 and 3-Bit Adders. In: Shakya S., Balsa V.E., Hoaxing W., Big Z. (eds) Proceedings of International Conference on Sustainable Expert Systems. Lecture Notes in Networks and Systems, vol 176. Springer, Singapore. https://doi.org/10.1007/978-981-33-4355-9_3

[22] M. Sakthimohan, J. Deny. (2020). An Enhanced 8x8 Vedic Multiplier Design by Applying Urdhva-Tiryakbhyam Sutra. International Journal of Advanced Science and Technology, 29(05), 3348 - 3358. http://sersc.org/journals/index.php/IJAST/article/view/12015

[23] S. M, E. R. G, P. Ramprasad, C. V. Reddy and B. S. Prasad, "Digital Fuel Monitoring System For Automobiles," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), 2021, pp. 6-10, doi: 10.1109/ICESC51422.2021.9532604.

[24] Hock Bing Lim, “A Soldier Health Monitoring System for Military Applications,” International Conference on Body Sensor Networks,2010.

.