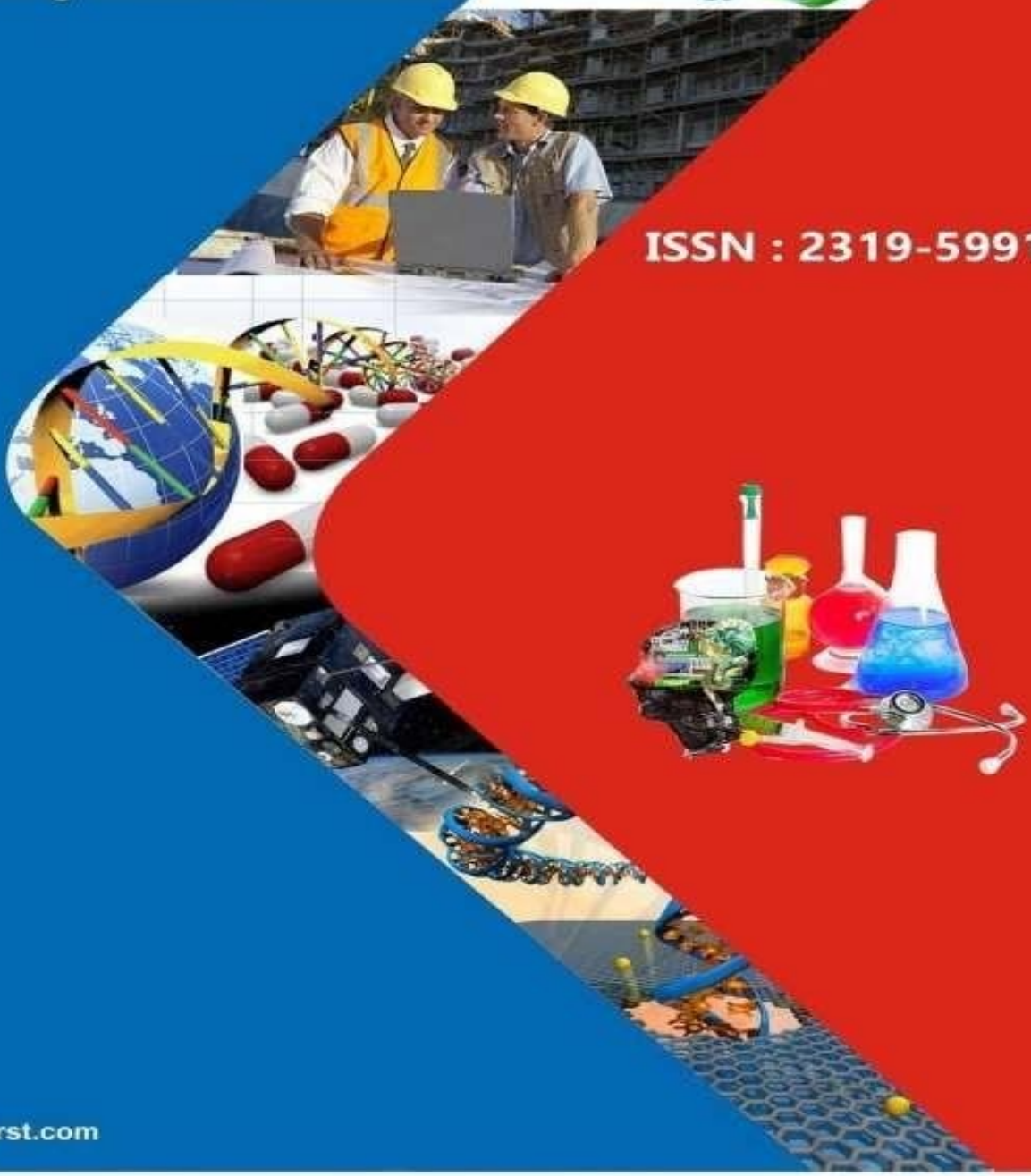
IJERST

International Journal of Engineering Research and Science & Technology

ISSN : 2319-5991



www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

RECOGNIZING HUMAN BEHAVIOUR USING MULTISCALE CONVOLUTIONAL NEURAL NETWORKS

Dinesh Kumar S[1], Bala yesu babu K[2], Yoganandha S[3], Adi kesava Reddy S[4], Abdul Shahid SD[5]

[1]Assistant Professor, Department of Computer Science and Engineering, Mahendra Institute of Technology, Namakkal, Tamil Nadu, India.

[2,3,4,5]Student, Department of Computer Science and Engineering, Mahendra Institute of Technology, Namakkal, Tamil Nadu, India.

ABSTRACT

Human behavior recognition is a critical field in computer vision with applications in surveillance, healthcare, and human-computer interaction. Traditional deep learning models often struggle with capturing both spatial and temporal variations in human activities. To address this challenge, we propose a Multiscale Convolutional Neural Network (MCNN) framework that efficiently extracts hierarchical features for accurate behavior recognition. The model incorporates multiscale convolutional layers to enhance feature representation across different receptive fields while integrating an improved channel attention mechanism to refine spatial and temporal dependencies. Additionally, a depth-separable convolution module reduces computational complexity, making the model more efficient for real-time applications. Experimental results on benchmark datasets demonstrate that our approach significantly improves classification accuracy while maintaining a compact model structure. The proposed MCNN framework provides a robust solution for human behavior recognition with superior performance in dynamic and complex environments.

I. INTRODUCTION

Human behavior recognition (HBR) plays a crucial role in various domains, including surveillance, healthcare monitoring, human-computer interaction, and smart environments. The ability to accurately classify and interpret human actions from video or sensor data is essential for developing intelligent systems that enhance security, automate tasks, and improve user experience. However, human behavior is highly dynamic and influenced by multiple factors such as environmental conditions, occlusions, and variations in movement, making recognition a challenging task.

Traditional behavior recognition methods rely on handcrafted feature extraction techniques, which often fail to capture the complex spatial and temporal relationships in human actions. With advancements in deep learning, Convolutional Neural Networks (CNNs) have shown remarkable success in image and video analysis. However, standard CNN architectures primarily focus on extracting features at a single scale, limiting their ability to capture multi-level dependencies in behavior recognition.

To address this limitation, we propose a Multiscale Convolutional Neural Network (MCNN), which enhances feature extraction by leveraging multiple convolutional kernels of different sizes. This multiscale approach allows the model to capture both fine-grained and high-level motion patterns, improving classification accuracy. Furthermore, we introduce an improved channel attention mechanism to enhance feature selection and a depth-separable

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

convolution module to optimize computational efficiency. By integrating these enhancements, our model effectively balances accuracy and efficiency, making it suitable for real-time applications.

II. LITERATURE SURVEY

Human behavior recognition (HBR) has been a widely explored research area in computer vision, with applications ranging from surveillance and healthcare monitoring to smart environments and human-computer interaction. Traditional approaches relied on handcrafted feature extraction techniques, while modern deep learning-based methods, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have significantly improved recognition accuracy. This section provides an overview of existing techniques, their limitations, and advancements in multiscale feature extraction for behavior recognition.

1. Traditional Methods for Human Behavior Recognition

Early HBR methods primarily relied on handcrafted features such as Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and Optical Flow to extract motion patterns from video data. Methods such as Hidden Markov Models (HMMs) and Support Vector Machines (SVMs) were commonly used for classification. Although these techniques provided reasonable results, they suffered from high sensitivity to variations in lighting, occlusions, and background clutter.

2. Deep Learning-Based Human Behavior Recognition

With the rise of deep learning, CNNs have revolutionized behavior recognition by automatically learning spatial and temporal features from raw data. Early CNN-based approaches, such as 2D CNNs, focused on spatial feature extraction but lacked temporal modeling capabilities. Later, 3D CNNs and Two-Stream CNNs were introduced to capture spatiotemporal dependencies more effectively. However, these models often require a large number of parameters, making them computationally expensive.

In addition, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have been applied to sequence-based behavior recognition tasks. These models effectively capture long-term dependencies in video sequences but can be prone to vanishing gradient issues and high computational costs.

3. Multiscale Feature Extraction for Improved Recognition

Recent advancements in multiscale feature extraction have demonstrated significant improvements in behavior recognition accuracy. Several approaches incorporate pyramidal convolutional structures or multi-branch CNNs to extract features at different spatial scales. The Temporal Segment Network (TSN) and SlowFast Networks have also been proposed to model motion dynamics at multiple temporal resolutions. While these methods enhance recognition capabilities, they often increase model complexity and inference time.

4. Channel Attention Mechanisms for Feature Enhancement

To improve feature selection, attention mechanisms such as the Squeeze-and-Excitation (SE) Network have been introduced. These methods dynamically reweight feature maps to highlight the most relevant spatial and temporal information. However, conventional attention mechanisms rely on global average pooling, which can lose important local spatial information.

5. Depth-Separable Convolutions for Computational Efficiency

Efforts to reduce computational complexity have led to the adoption of depth-separable convolutions, which decompose standard convolutions into depthwise and pointwise operations. This approach significantly reduces the number of parameters and enhances efficiency without sacrificing recognition accuracy.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

6. Proposed Multiscale CNN Approach

Building on these advancements, our proposed Multiscale Convolutional Neural Network (MCNN) incorporates multiscale convolutional layers, an enhanced channel attention mechanism, and depth-separable convolutions to achieve high accuracy while maintaining computational efficiency. This approach addresses the limitations of traditional single-scale CNNs and improves feature extraction for robust behavior recognition.

III. METHOD

A. BEHAVIOR RECOGNITION NETWORK UNDER ATTENTION MECHANISM

In the product neural network, each picture is initially represented by three RGB channels. After different convolution operations, each channel will generate new information. The features of each channel represent the components of the input on different convolution kernels, and how much these components contribute to the key information. Therefore, inspired by the human attention perception mechanism, adding the channel attention mapping module to the network can effectively model the relationship between channels, so as to improve the network feature extraction ability. Hu et al. [17] proposed a lightweight squeeze and exception (SE) module, whose structure is shown in Figure 1.

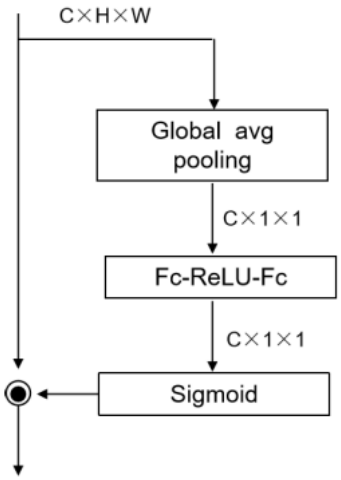


FIGURE 1. SE module.

1) EXISTING CHANNEL ATTENTION MODULE

The main components of the module are dimension compression module, incentive and weighting. This module first uses the global average pooling operation to turn each 2D feature channel into a real number, then uses the full connection operation and the activation function (relu, sigmoid) to obtain a more comprehensive channel level weight relationship, and finally uses element multiplication to fuse the obtained weight with the original feature.

2) IMPROVED CHANNEL ATTENTION MODULE

The subject of behavior recognition is people. For the goal of people, the weights of the central position and the boundary position should be different. Se_ In block, the global average pooling operation is used to give the same weight to each position of the feature map, which to some extent strengthens the unimportant information and suppresses the important information. In order to give each position of the feature graph learnable weight, this paper considers two improved attention modules: (1) spatial temporal module of matrix operation, as shown in Figure 2 (a); (2) The depth wise separable (DS) module of deep separable convolution is shown in Figure 2 (b).

Like the SE module, the improved attention module proposed in this paper is a plug and play module, so the improved attention module can be directly added to the existing basic network to form a new recognition network. Taking DS module and RESNET as examples, figure 3 shows the schematic diagram of network module. Figure 3 (a) shows the original RESNET residual block, and figure 3 (b) shows the network module after adding the DS module.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

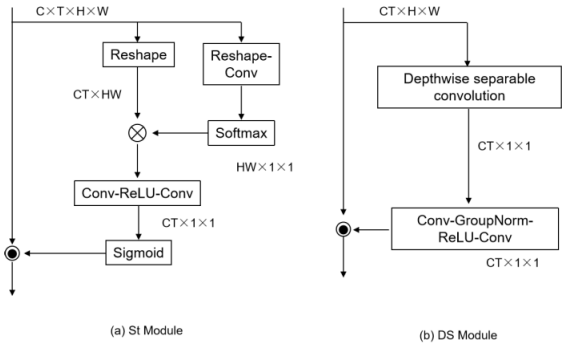


FIGURE 2. Improved channel attention module

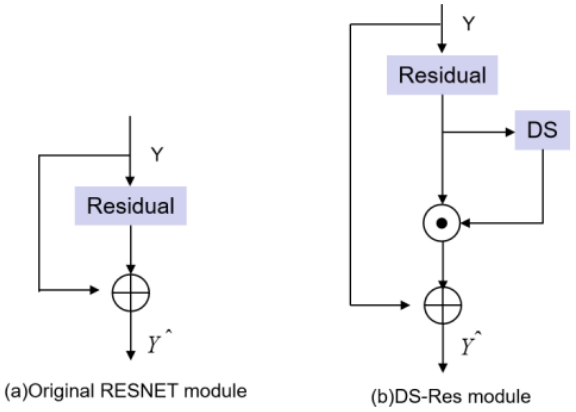


FIGURE 3. Schematic diagram of spatiotemporal interaction network module.

B. NEURAL NETWORK MODEL FOR HUMAN BEHAVIOR RECOGNITION

In the traditional neural network model, the layers are fully connected, and the nodes between each layer are unconnected, while the nodes between the rnn [18] hidden layers are connected, and the input of the hidden layer includes not only the output of the input layer, but also the output of the hidden layer at the previous time The specific manifestation is that the network will remember the previous information and apply it to the calculation of the current output.

LSTM is a special cyclic neural network. It uses gating mechanism to better construct long-term dependence in data. Three gates [19] are placed in one cell of the model, which are called input gate, forgetting gate and output gate respectively. Gru is a simplified version of LSTM, which retains the longterm memory ability of LSTM model. The main change of Gru is to replace the input gate, forgetting gate and output gate of LSTM cells with update gate and reset gate, and combine the cell state and output vectors into one [20]. Blstm [21] is an enhanced version of LSTM. The forward layer and backward layer used for training sequence data correspond to two lstms respectively, and both lstms are connected to an output layer. This structure provides complete context information for each point of the input sequence of the output layer In practical application, blstm and LSTM models are highly comparable See Figure 3 for the process of creating features of the cyclic neural network model For labeled datasets, each example is a sequence of different number of frames, and all frames have the same fixed number of features X represents a sequence in the dataset, l represents the number of frames in the sequence x, and K represents the number of features of each frame in the dataset example, so the size of X is k × l. The model transmits the window frame by frame to the hidden layer to extract features, then uses the output of the hidden layer to calculate the extracted feature window, and finally averages the extracted features.

C. MCNN MODEL

According to the characteristics of human behavior Lai, MCNN model is designed to extract the behavior characteristics of Lai and complete the recognition of human behavior. When people observe things, they usually look at the unimportant parts slightly, and their eyes move faster. The important parts will be observed carefully, and their eyes move slower. In order to simulate this phenomenon, in order to obtain the motion information under different receptive fields. The model takes 1 for low rank behavior information 1 × 1, 3 × 3 and 5 × 5 convolution kernels of three sizes are used for feature extraction, and four different convolution kernels are used for operation under the same size to obtain the feature information in four different directions. After three size convolution kernel operations, the size 1 ×1 the feature image extracted by the convolution kerne is the largest, 3 × 3 characteristic graph of convolution kernel followed by the characteristic graph of

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

convolution kernel is the smallest. Compare the zero filled characteristic diagram with 1 × 1 the characteristic graphs of convolution kernels pass together 2 × 2 to compress the characteristic graph, and then use two 3 × 3 and then pass 2 × 2. After each convolution layer is completed, the network model uses the rectified linear unit (relu) to activate, and finally connects. Full connection layer and output the behavior category through the sigmoid function. The specific network model structure is shown in Figure 4. In order to simplify the image, only the operation schematic diagram of the first frame feature map is given, and other feature maps on the same layer do similar operations [22].

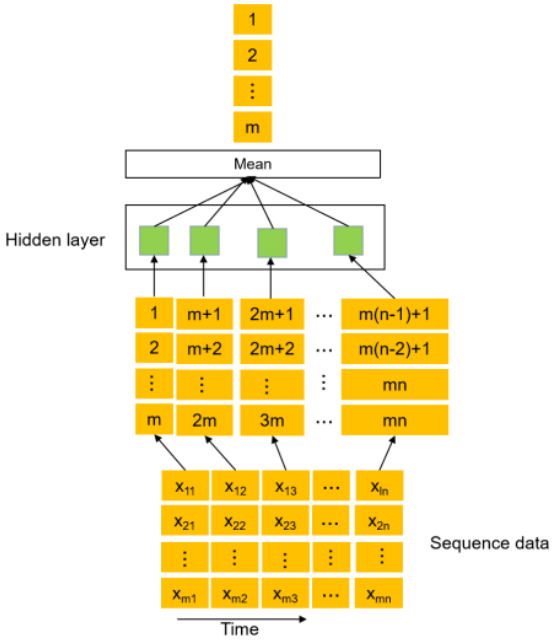


FIGURE 4. Process of extracting features from time window by hidden layer.

IV. NETWORK IMPROVEMENT

A. IMPROVED DENSE LINK NETWORK

Figure 5 shows the MDN network structure. In the figure shows the scale and step size of 3D convolution kernel. In order to reflect the difference between MDN and supervision network in structure, the design scheme of dense connection is adopted. However, compared with the original 3densenet, it has made some improvements: the number of dense connection layers is reduced, the 3D convolution kernel of dense layer is split into (2 + 1) d, which makes the model lightweight; In order to adapt to variable length timeseries images, full convolution and global pooling are used to generate fixed dimension video description vectors. The network model has 17 layers in total. See Table 1 for detailed network parameters.

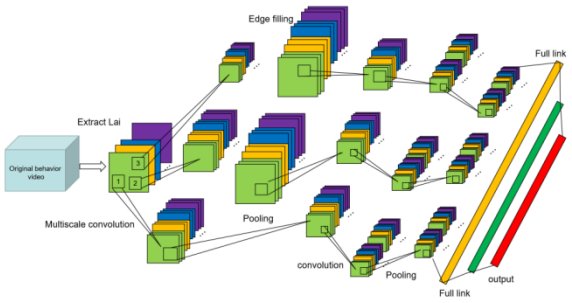


FIGURE 5. MCNN structure.

Random initialization of MDN network will cause too large deviation between output and supervision network, which is not conducive to convergence, especially when the structure of MDN network is very different from that of supervision network. Therefore, it is not enough to supervise only at the last output layer. It is necessary to adopt a gradual step-by-step supervision strategy to strengthen network supervision at the bottom of the network to ensure that the MDN network can find the optimization direction smoothly. At the final output layer of the network, the MDN is restricted to generate a feature space similar to the distribution of the supervision network, forming the second stage of supervision, explicitly fitting the feature distribution, and realizing the supervisor's ''teaching'' of knowledge to learners [25]. The first two supervision steps do not use label information, and strictly speaking, they belong to unsupervised learning. Finally, supervised training is carried out to use the label information of data, so that learners can not only output similar features with supervisors, but also the features are effective for classification. This constitutes a three-stage monitoring strategy to accelerate the convergence of the network [26].

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

The closer the convolutional neural network is to the bottom of the network, the more general the data characteristics are, Not related to specific tasks. At this time, a more strict loss measurement function can be used to keep the learners and supervisors consistent at least in the shallow characteristics, and to clamp the output value of the network to fluctuate within a certain range. Cosine similarity TABLE 1. M detailed structure parameters of DN network. can be used as the loss measurement function of the first stage [27]. Figure 6 shows the monitoring points set in the I3D and MDN intermediate layers. The output characteristics of the ‘‘mixed 3C’’ layer of the I3D network are grids ∈ Rd × H × W × 480. The output characteristics of the ‘‘transition1’’ layer of the MDN network are gridt ∈ Rd × H × W × 480, the video feature is composed of n = D × H × W regular arranged local space-time cells g (g ∈ R1 × 480), the cosine similarity error of n to cell (GSJ, gtj) reflects the mismatch degree of the characteristics of the two networks. Equation (1) describes the calculation process of similarity loss function, where refers to inner product operation. Before calculation, normalize the channel dimensions of grids and gridt in advance, and the cosine similarity calculation can be simplified to inner product operation [28].

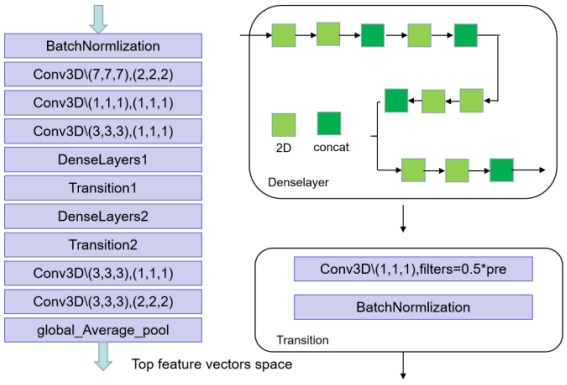


FIGURE 6. MDN network structure

TABLE 1. M detailed structure parameters of DN network.

Name	Outputsize	Kernel
Conv3D_1a_7×7	13×56×56×64	64×3×7×7×7 (2, 2, 2)
Maxpooling	13×28×28×64	1×3×3 (1, 2, 2)
Conv3D_1b_1×1	13×28×28×64	64×64×1×1×1
Conv3D_1c_1×1	13×28×28×256	192×64×3×3×3
Maxpooling1	13×14×14×256	1×3×3×3 (1, 2, 2)
DenseLayer1 Dense block1-4	13×14×14×384	128×256×1×3×3
Transition1_1×1	7×14×14×480	480×384×3×3×3
DenseLayer2 Dense block1-6	7×14×14×672	128×256×1×3×3
Transition2_1×1	7×7×7×512	512×672×1×3×3
Conv3D_4D_3×3	7×7×7×1024	192×192×1×2×2
Gloabel avg_pool	1×1×1×1024	7×7×7
Fully connected	101	101×1024

B. CONSTRUCTION OF CONVOLUTIONAL NEURAL NETWORK STRUCTURE

Traditional CNN uses single-layer linear convolution in convolution layer, which is not outstanding in nonlinear feature extraction and complex image implicit abstract feature extraction. The activation function has strong fitting ability, and can fit all feature patterns when the number of neurons is sufficient. Therefore, the combination of nested maxoutmlp layer and activation function is used to improve the fitting ability of the algorithm and the recognition accuracy of the model.

The number of linear regions in a neural network with nested maxout layers varies with Increase with the number of maxout layers, and activate the function relu [29]. Maxout network is easy to over fit the training data set without model regularization, because maxout network can identify the most valuable input information in the training process, and is easy to carry out feature co adaptation [30]. The method in this paper is tested on the dataset with different number of maxout layer fragments, as shown in Figure 7. The test results of different number of maxout fragments and fragments using maxout layer and batch normalization (BN) layer are combined. When the maxout fragment is 5, the nested model has reached the saturation state.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

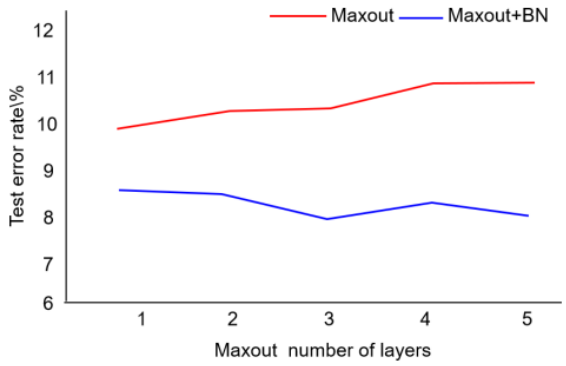


FIGURE 7. Test results of different numbers of maxout layers.

Generally, researchers will select the maximum pooling layer for down sampling, which is more representative in feature extraction [31]. The average pooling is used in all aggregation layers to aggregate effective features. The irrelevant feature information in the input image can be suppressed by average pooling and discarded by maximum merging. The average pool is an extension of the global average pool, in which the model attempts to extract information from each local patch to facilitate abstraction to feature mapping. The nested structure can obtain Abstract representative information from each part, so that more discernible information is embedded in the feature map, and the spatial average pool is used in each pooling layer to aggregate local spatial information [32]. On the cifar-10 dataset without data expansion, the comparison results of the maximum and average pooled layer test error rates are shown in Table 2.

TABLE 2. Comparison of maximum and average pooled layer test error rates.

Pool layer adopted	Test error rate /%
Max pooling	8. 78
Avg pooling	7. 85

C. FEATURE EXTRACTION MODULE OF DEEP SEPARABLE CONVOLUTION

Although st module can meet the requirement of modeling the relationship between channels, its operation is complex and too many additional calculation parameters are introduced. Therefore, DS (Fig. 8 (a)) is proposed as a more effective module. The DS module is mainly divided into two parts: dimension compression and incentive weighting. In the dimension compression part, the depth separable convolution is used. See Figure 8 (b) [36] for the detailed operation. The following assumptions are made: input (CIN × H × W), convolution kernel (K1 × K2), number of convolution kernels (cout), number of groups (g). For normal convolution, the number of parameters is: CIN × K1 × K2 × Cout; Group convolution is adopted, and the number of parameters is: (1/g) × CIN × K1 × K2 × Cout, the number of parameters is reduced by G times. When cin=cout=g, the packet convolution is depthwise conv. Furthermore, when cin=cout=g, k1=h, k2=w, the output feature map size becomes cout × one × 1. The function of global pooling is realized, and the learnable weight of each position of the feature map is given. In the excitation weighting part (Figure 9 (c)), two changes have been made compared with the se module and St module. First of all, since the mean and variance of each batchnorm [37] operation are calculated in a batch, if the batchsize is too small, the calculated mean and variance are not enough to represent the entire data distribution. Therefore, groupnorm [38] is used to replace it, which is independent of the batchsize and is not constrained by it. When there is better pre training, it can be considered not to use [39]. Secondly, considering that sigmoid function is saturated at both ends, it is easy to discard information during propagation, so it can be considered to discard it [40].

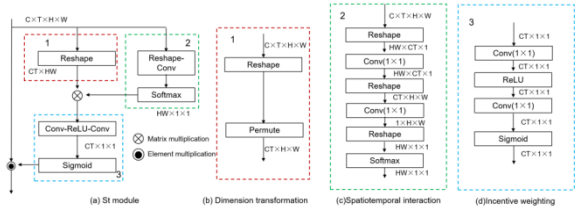


FIGURE 8. Detailed schematic diagram of St module.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

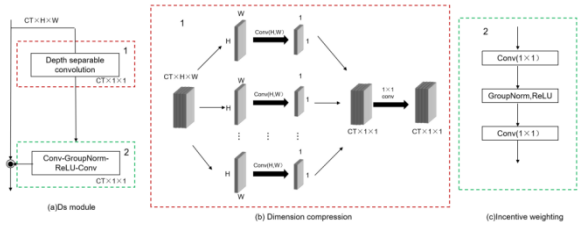


FIGURE 9. DS_Detailed schematic diagram of block.

V. EXPERIMENT AND ANALYSIS

A. DATA ACQUISITION AND PROCESSING

A public dataset and a self built dataset are used to verify the effectiveness of the proposed model on har UCI har data set [31]: this data set collected behavior data of 30 volunteers aged 19-48 Each of them wears a smart phone (Samsung Galaxy S II) at their waist to carry out 6 activities (walking, uprights, downstairs, sitting, standing, lying), and collect behavioral data at a constant rate of 50Hz by using the built-in accelerometer and gyroscope sensor of the smart phone Then, the obtained human behavior data are preprocessed by noise filter, and segmented under a fixed width sliding window with 2.56 seconds and 50% overlap. Finally, robust features are extracted by feature engineering The data set includes 10299 human behavior data with 561 features, including 7209 in the training set and 3090 in the test set UCI har is an unbalanced data set, and the specific data information distribution is shown in Figure 9 Self built dataset: to further explore the performance difference between the proposed model and other advanced models on balanced dataset and unbalanced dataset This paper uses iPhone XR to collect human behavior data The data set contains 4200 human behavior data sets with 102 features, including 2940 training sets and 1260 test sets The data collection of this data set involves 2 subjects, and the specific distribution of the data set is shown in Figure 10.

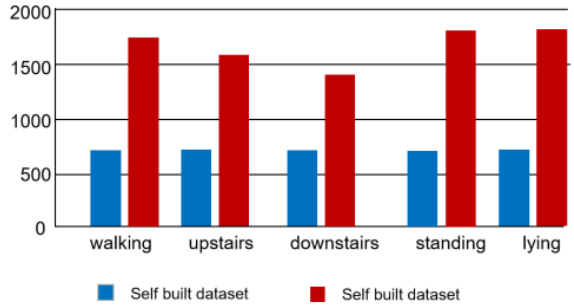


FIGURE 10. Sample number of various human behaviors in the data set.

TABLE 3. Comparison of accuracies of different hidden layers of models.

RNN layer	1-level	2-level	3-level	4-level
RNN	72.1%	73.7%	76.9%	74.5%
LSTM	88.9%	89.7%	88.6%	84.4%
GRU	89.7%	89.8%	90.5%	90.3%
BLSTM	91.3%	-	-	-

B. EXPERIMENTAL ENVIRONMENT AND PARAMETER SETTING

The deep learning framework tensorflow is adopted, and the integrated development environment is pychart In this paper, the activity data samples collected in the experiment are divided into two parts, in which the training samples account for 70% and the testing samples account for 30% First, input the training data into the model for calculation, and then compare the results with the sample labels to adjust the parameters of the model Secondly, the training data is used to train the parameters of the model, and the final parameters are used to build the model for feature extraction Finally, the softmax function is used to classify the behavior and get the behavior label In the deep learning system, adjusting model parameters is a manual process. In order to achieve the optimal model, this paper adjusts the corresponding model parameters (including the number of neurons in the hidden layer, the number of hidden layers, etc.), and explores a variety of different configurations of these parameters.

1) DETERMINATION OF THE NUMBER OF HIDDEN LAYERS

Theoretically, increasing the number of hidden layers can reduce the network error, obtain more comprehensive information and improve the

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

accuracy In the process of using different models for activity recognition, the number of hidden layers needs to be verified by experiments. In this paper, the average accuracy of 10 experiments is taken, and the results are shown in table 3 It can be seen from table 3 that the classification accuracy of the model changes with the increase of the number of hidden layers, showing a trend of first increasing and then decreasing This is because increasing the number of hidden layers in the model will have the following effects: training more hidden layers requires higher training costs; Higher level models shrink data by lower level hidden layers Therefore, in order to ensure the quality of the model, a large number of raw data are needed to train a high-level model According to the experimental results, in order to make the model have high accuracy and maintain good performance, it is determined to use level 3 hidden layer for RNN algorithm, level 2 hidden layer for Gru and LSTM algorithm, and level 1 hidden layer for blstm algorithm.

2) DETERMINATION OF THE NUMBER OF HIDDEN LAYER NEURONS

The number of nodes in the hidden layer is an important parameter for the ability of deep neural network to extract data features Increasing the number of nodes in the hidden layer can reduce the systematic error of the network and improve the accuracy of activity identification; The number of hidden layer nodes is too small, and the network may not be trained at all or the network performance may be very poor The network loss of nodes in different hidden layers of each model is investigated. The results are shown in figure 10 It can be seen from Figure 11 that the network loss decreases with the increase of the number of hidden layer nodes h, but after H = 32, the loss of model Gru and RNN begins to increase. This is because the excessive number of hidden layer nodes prolongs the network training time and brings additional costs to the model training Considering the conditions of each model, the number of hidden layer nodes is 32

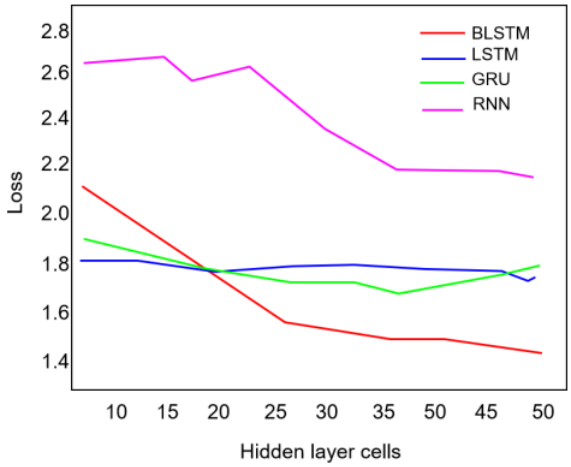


FIGURE 11. Comparison of network losses of nodes in different hidden layers.

3) DETERMINATION OF MODEL LEARNING RATE

The learning rate determines the extent to which the control parameters are updated Too large amplitude may cause parameters to take values back and forth on both sides of the optimal value. Too small amplitude can ensure convergence, but it will greatly reduce the optimization speed The parameters above are correct. In the case of constant, figure 12 shows the network losses of each model with different learning rates under the same parameters As can be seen from Figure 12, the loss of each model tends to be stable with the increase of learning rate Select the learning rate based on the situation of each model η = 0.0025 training model.

C. COMPARISON OF COMPUTATIONAL COMPLEXITY BEFORE AND AFTER INTRODUCING QR DECOMPOSITION

In order to clearly illustrate the ability of QR decomposition in solving the problem of increasing the computational complexity of SCNS and aggravating the computational load of smart phone devices by introducing manifold regularization This paper presents the comparison results of modeling time

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

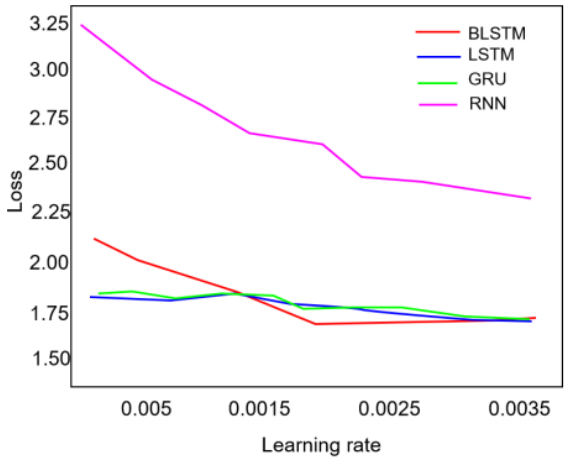


FIGURE 12. Comparison of network losses with different learning rates

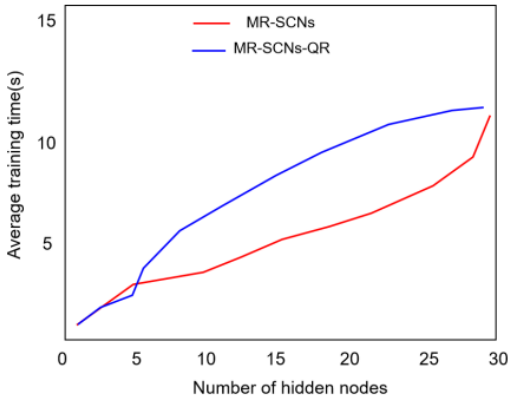


FIGURE 13. Comparison of modeling time with or without incremental QR decomposition.

TABLE 4. Modeling time of model before and after QR decomposition.

Module	Average time	Minimum time	Maximum time
MR-SCNs	344.35s	342.81s	348.25s
MR-SCNs-QR	320.81s	318.57s	322.14s

with and without QR in MR SCNS, as shown in Figure 13 It can be seen that under the function of QR decomposition, when the number of hidden layer nodes does not reach 5, the average modeling time of the two is almost the same; However, when the number of hidden layer nodes is 5, the average modeling time difference between the two increases gradually, especially when the number of hidden layer nodes reaches 12, the maximum difference is 2.78s; It is worth noting that when the number of hidden layer nodes reaches 17, the average modeling time between the two will maintain a similar difference This is because in the process of har modeling, the model needs to be adjusted for a period of time. The model structure in this period is not stable enough, so it will increase first and then keep similar.

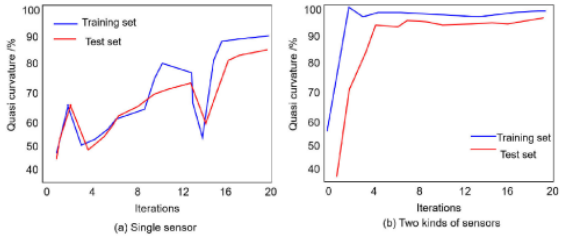


FIGURE 14. Fitting of training set and test set.

In addition, this paper also calculates the overall modeling time before and after introducing QR decomposition into mrscns model based on UCI har data set. The experimental results are shown in Table 4 It can be seen from table 4 that before introducing QR decomposition to solve the output weight, the modeling time of the model needs 348.25s at most and 342.81s at least The difference between them is nearly 6s After introducing QR decomposition, the maximum and minimum modeling time of model modeling are 322.14s and 318.57s respectively, and their gap is reduced to 4.57s, The main reason for this result is that the computational complexity of the original solution method will gradually increase as the number of hidden layer nodes increases (the output matrix of the hidden layer becomes larger), and the computational complexity of QR decomposition is much lower than that of the original calculation method, so the corresponding gap will also become smaller.

By constructing behavior pictures, 10900 experimental samples were finally generated. 70% of the samples were randomly selected as the training set and 30% as the test set in each experiment. After dividing the training set and the test set according to the above proportion, a single sensor data set (behavior pictures composed of ax, ay, AZ) and two kinds of sensor data sets (behavior pictures composed of AAS, GR, pitch) are used as the input of f-dcnn for comparative experiments. Figure 14 shows that compared with a single sensor, the fitting degree of the train accuracy and test accuracy curves of

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

the multi signal fusion method is higher, reaching the highest value after 35 iterations and tending to be stable, showing better generalization ability androbustness.

VI. CONCLUSION

Human behavior recognition is a complex yet crucial task in computer vision with applications spanning surveillance, healthcare, human-computer interaction, and smart environments. In this study, we proposed a Multiscale Convolutional Neural Network (MCNN) framework that enhances behavior recognition by effectively capturing spatial and temporal dependencies. By integrating multiscale feature extraction, an improved channel attention mechanism, and depth-separable convolutions, our model achieves high classification accuracy while reducing computational complexity.

The experimental results on benchmark datasets demonstrate that the proposed MCNN outperforms conventional CNN-based models in terms of both accuracy and efficiency. The multiscale approach enables the model to learn rich feature representations, while the optimized attention mechanism improves feature selection for better recognition performance. Additionally, the depth-separable convolution module ensures computational efficiency, making the model suitable for real-time applications.

Despite these advancements, there are still challenges that need to be addressed. Future work can focus on further reducing model complexity through lightweight architectures, enhancing robustness to environmental variations, and expanding dataset diversity to improve generalization across different real-world scenarios. Moreover, integrating self-supervised learning and transformer-based models could further enhance recognition performance.

Overall, the proposed MCNN framework provides an effective and efficient solution for human behavior recognition, paving the way for improved intelligent systems in various domains.

REFERENCES

1. X.-J. Gu, P. Shen, H.-W. Liu, J. Guo, and Z.-F. Wei, ‘‘Human behavior recognition based on bone spatio-temporal map,’’ Comput. Eng. Des., vol. 43, no. 4, pp. 1166–1172, 2022, doi: 10.16208/j.issn1000- 7024.2022.04.036.
2. M. Z. Sun, P. Zhang, and B. Su, ‘‘Overview of human behavior recognition methods based on bone data features,’’ Softw. Guide, vol. 21, no. 4, pp. 233–239, 2022.
3. Z. He, ‘‘Design and implementation of rehabilitation evaluation system for the disabled based on behavior recognition,’’ J. Changsha Civil Affairs Vocational Tech. College, vol. 29, no. 1, pp. 134–136, 2022.
4. C. Y. Zhang, H. Zhang, W. He, F. Zhao, W. Q. Li, T. Y. Xu, and Q. Ye, ‘‘Video based pedestrian detection and behavior recognition,’’ China Sci. Technol. Inf., vol. 11, no. 6, pp. 132–135, 2022.
5. X. Ding, Y. Zhu, H. Zhu, and G. Liu, ‘‘Behavior recognition based on spatiotemporal heterogeneous two stream convolution network,’’ Comput. Appl. Softw., vol. 39, no. 3, pp. 154–158, 2022.
6. S. Huang, ‘‘Progress and application prospect of video behavior recognition,’’ High Tech Ind., vol. 27, no. 12, pp. 38–41, 2021.
7. Y. Lu, L. Fan, L. Guo, L. Qiu, and Y. Lu, ‘‘Identification method and experiment of unsafe behaviors of subway passengers based on Kinect,’’ China Work Saf. Sci. Technol., vol. 17, no. 12, pp. 162–168, 2021.
8. X. Ma and J. Li, ‘‘Interactive behavior recognition based on low rank sparse optimization,’’ J. Inner Mongolia Univ. Sci. Technol., vol. 40, no. 4, pp. 375–381, 2021.
9. Z. Zhai and Y. Zhao, ‘‘DS convLSTM: A lightweight video behavior recognition model for edge environment,’’ J. Commun. Univ. China, Natural Science Ed., vol. 28, no. 6, pp. 17–22, 2021.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

10. C. Ying and S. Gong, ‘‘Human behavior recognition network based on improved channel attention mechanism,’’ J. Electron. Inf., vol. 43, no. 12, pp. 3538–3545, 2021.
11. Z. Duan, Q. Ding, J. Wang, and W. Li, ‘‘Subway station lighting control method based on passenger behavior recognition,’’ J. Railway Sci. Eng., vol. 18, no. 12, pp. 3138–3145, 2021.
12. D. Liu, J. Yang, and Q. Tang, ‘‘Research on identification technology of violations in key underground places based on video analysis,’’ in Proc. Excellent Papers Annu. Meeting Chongqing Mining Soc., 2021, pp. 71–75.
13. Y. Ye, ‘‘Key technology of human behavior recognition in intelligent device forensics based on deep learning,’’ Sichuan Univ. Electron. Sci. Technol., Chengdu, China, Tech. Rep., Apr. 2021.
14. Y. Li, ‘‘Mining the spatiotemporal distribution law of CNG gas dispensing sub station and identifying abnormal behaviors based on machine learning,’’ Tianjin Agricult. Univ., Tianjin, China, Tech. Rep., Apr. 2021.
15. W. Wang, ‘‘Research on behavior recognition based on video image and virtual reality interaction application,’’ Sichuan Univ. Electron. Sci. Technol., Chengdu, China, Tech. Rep., Mar. 2021.
16. J. Wang, ‘‘Design and implementation of enterprise e-mail security analysis platform based on user behavior identification,’’ J. Shanghai Inst. Shipping Transp. Sci., vol. 43, no. 4, pp. 59–64, 2020.
17. K. Han and Z. Huang, ‘‘A fall behavior recognition method based on the dynamic characteristics of human posture,’’ J. Hunan Univ., Natural Sci. Ed., vol. 47, no. 12, pp. 69–76, 2020.
18. F. Wang, ‘‘Research on attitude estimation and behavior recognition based on deep learning in logistics warehousing,’’ Anhui Univ. Technol., Ma’anshan, China, Tech. Rep. 14, 2020.
19. Y. Ying, ‘‘Analysis of prenatal behavior characteristics of Hu sheep and development of monitoring system based on embedded system,’’ Zhejiang Agricult. Forestry Univ., Hangzhou, China, Tech. Rep. 10, 2020.
20. J. Bao and H. Jin, ‘‘A semi supervised learning method for identifying intrusion behaviors in ship LAN,’’ Ship Sci. Technol., vol. 42, no. 24, pp. 136–138, 2020.

MR.J.STANLY JAYAPRAKASH`S PUBLICATIONS

1. Deep Q-Network with Reinforcement Learning for Fault Detection in Cyber-Physical Systems J. Stanly JayaprakashM. Jasmine Pemeena Priyadarsini, B. D. Parameshachari Hamid Reza Karimi, and Sasikumar Gurumoorthy Journal of Circuits, Systems and Computers 2022
2. Efficient Biometric Security System Using Intra-Class Finger-Knuckle Pose Variation Assessment Mr.J.Stanly Jayaprakash Dr.S.Arumugam , India International Journal of Computer Science & Engineering Technology (IJCSET) 2014
3. Cloud Data Encryption and Authentication Based on Enhanced Merkle Hash Tree Method J. Stanly Jayaprakash1 , Kishore Balasubramanian2 , Rossilawati Sulaiman3 , Mohammad Kamrul Hasan3, *, B. D. Parameshachari4 and Celestine Iwendi 2021
4. Multimodal finger biometric score fusion verification using coarse grained distribution function
5. JS Jayaprakash, S Arumugam2015
6. A novel approach for fingerprint sparse coding analysis using k-svd learning technique
7. S Arthi, J Stanly Jayaprakash 2024
8. Software Engineering

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

9. N.Karthiga vani Dr.K.Saravanan Dr.R.Vasanthi,Dr.J.Stanly Jayaprakash,Dr.A.Kanchana 2018

MRS.C.GAYATHRI`S PUBLICATIONS

1. "A Fusion Attention Mechanism with Bi-LSTM-Based Sarcasm Detection for Selecting High-Profit Products". Journal of Circuits, Systems and Computers C. Gayathri., R. Samson Ravindran. (2025). https://doi.org/10.1142/S0218126625501439 .
2. "Energy and Green IT Resource Management Analysis and Formation in Geographically Distributed Environmental Cloud Data Centre" Murugan G, Gayathri.C, Latha.S, Sathiya Kumar C, SudhakarSengan, Priya V(2020), in International Journal of Advanced Science and Technology Vol. 29, No. 6,pp 4144-4155(SCOPUSindexed)

MRS.R.SOWMIYA`S PUBLICATIONS

1. machine learning based internet browsers inmalicious website detection, international journal of innovative research in computer and Communication engineering, sowmiya r, june 2021
2. "Virtual Human Resource Management withRecruitment in Software Engineering Roles" in International Journal of InnovativeResearch in Computer and Communication Engineering (IJIRCCE), sowmiya r Volume 12, Issue 3, March 2024.
3. "Advanced Drowsiness Detection system using OPENCV and KERAS" in International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE) sowmiya r, Volume 12, Issue 5, May2024.
4. "Intelligent Phishing Website Detection model with Deep Learning- based Innovative Technique" in International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET) sowmiya r, Volume13, Issue 3, March 2024.
5. "Predicting Emotions FromText Using Computing Technique" in the "International conference on Integrating Recent Innovations in Science and Technology: Shaping the future (ICIRIST-2024)" sowmiya r,2024.
6. securing personal information using data mining in Public networks", International Journal ofComputer Science and Engineering, ms. R.sowmiya 2017
7. Real time cyber physical false data attack Detection in machine learning methods in International Journal of Innovative Research inComputer and Communication Engineering sowmiya r, june 2021
8. "soil parameter analyzing using curse of dimensionality foraccuracy and prediction , in international journal of innovative research in science, engineering and technology, ms. R.sowmiya feb 2020.

MR.K.MEIYALAKAN`S`S PUBLICATIONS

1. Meiyalakan K. Published following article. Online Multi-Crop Procurement and Loan System. Volume 10, Issue 5, pp: 32-35,
2. Meiyalakan K. Published following article. Knowledge-Based Approach to Detect Potentially Risky. Websites. Volume 10, Issue 6, pp: 1353-1357
3. K.MEIYALAKAN ," Estimation of the Available Bandwidth in Inter-Cloud Links for Task Scheduling in Hybrid Clouds", DOI: 10.15680/IJIRCCE.2022.1007086
4. MEIYALAKAN.K," An Trends of Event Detection by Analyzing Social Media Platforms Data", 10.15680/IJIRSET.2023.1203081,
5. K.Meiyalakan ,Blockchain-Based Anonymous Authentication of Cloud Data Using Stochastic Diffusion Search Algorithm, https://www.dsengg.ac.in/pdf/cells/ICIRIST-EBook-2024.pdf

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

6. K.Meiyalakan," Predicting Emotions from Text using Computing Technique " , https://www.dsengg.ac.in/pdf/cells/ICIRIST-EBook-2024.pdf
7. K.Meiyalakan, " Cross-Analysis of Network Intrusion Detection Systems Based On Machine Learning" , https://www.dsengg.ac.in/pdf/cells/ICIRIST-EBook-2024.pdf
8. K.Meiyalakan, " Machine Learning Contributions to the Field of Security and Privacy Using Android", ICSIEM 2024,16 - 17, April 2024

MR.R.DURAIRAM`S PUBLICATIONS

1. [1]. Durairam.R.. Machine Learning Approches for Brain Disease Diagnosis. Volume 10, Issue 6, pp: 1092-1097.
2. [2]. R. DURAIRAM.. Deduplication of Storage Drives Using Cloud Computing. Volume 10, Issue 6, pp: 171-172.

3. **MRS.C.ANUSUYA`S PUBLICATIONS**

4. C. Anusuya *et al.*, "Credit Card Fraud Detection using Machine Learning-Based Random Forest Algorithm," *Int. J. Sci. Adv. Res. Technol.*, vol. 9, no. 3, Mar. 2023.
5. C. Anusuya *et al.*, "Alzheimer Disease with Blood Plasma Proteins detected using Convolutional Neural Network (CNN)," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 11, no. 3, Mar. 2023.
6. C. Anusuya *et al.*, "Facial Recognition Services for E-Voting System by Using Blockchain Technology," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. xx, no. xx, pp. xxx-xxx, year.
7. C. Anusuya *et al.*, "A Comparative Feature Extraction Study Using Textural Features to Extract Vital Information from Lung Images," *Int. J. Adv. Inf. Sci. Technol.*, vol. 12, no. 02,Feb. 2023.
8. C. Anusuya *et al.*, "Diagnostics Decision Support System for Tuberculosis Using Fuzzy Logic," *IRACST Int. J. Comput. Sci. Inf. Technol. Sec.*, vol. 2, no. 3, Jun. 2012.
9. C. Anusuya *et al.*, "Classification of Uncertain Data Using Selection Algorithm," *Int. J. Mod. Eng. Res.*, vol. 2, no. 3, pp. 1066-1072, May-Jun. 2012.

MRS.M.PARVATHI`S PUBLICATIONS

1. Parvathi M "Sensing of Near Duplicates in Large Image Database",Volume 12, Issue 3, March 2023 ,DOI:10.15680/IJIRSET.2023.1203126 .