Development of Image Recognition Teaching Model by the AR Technology

Mangalampalli Sesha Sai Lakshmi Lavanya[1], Botcha Kishore Kumar[2]

[1,2]Assistant Professor, CSE (Data Science) Department CMR Engineering College, Hyderabad.

Abstract: The technology known as augmented reality, or AR, has shown great promise in a number of industries, including education. This research addresses the construction of an image recognition teaching paradigm leveraging AR technology. The incorporation of AR into educational settings offers unique opportunity to enhance learning experiences by offering immersive, interactive, and engaging information. By merging image recognition algorithms with AR, educators may design novel teaching models that adapt to varied learning styles and stimulate active involvement. This study explores the design, implementation, and potential effect of an AR-based image recognition teaching paradigm, highlighting its merits and problems in the educational setting.

Keywords: Augmented Reality, Image Recognition, Teaching Model, Education, Technology Integration.

I.ITNRODUCITON

With its innovative combination of digital augmentation and real-world interaction, augmented reality (AR) is at the forefront of technological innovation, going beyond conventional computer paradigms. In contrast to virtual reality, which creates wholly artificial worlds, augmented reality (AR) modifies the actual world by superimposing digital data and virtual objects, so fusing the virtual and the real. This revolutionary technology has impacted many areas of contemporary life, including gaming, entertainment, healthcare, education, and business applications. It has completely changed how we view and engage with our environment [1].

The development of AugmentAed Reality (AR) has been propelled by advancements in Artificial Intelligence (AI), which are essential in augmenting the functionalities and user experiences of AR systems. The core of augmented reality (AR) technology is AI algorithms, which allow for seamless interaction between digital information and the actual world as well as real-time interpretation of the surroundings. AR systems are able to recognize and track objects, comprehend spatial connections, and provide individualized experiences for each user by utilizing machine learning and computer vision [5.

The adaptability of AR across a variety of disciplines is one of the primary factors promoting its wider adoption. For example, AR in healthcare gives surgeons better visualization tools so they can precisely plan and execute surgeries by superimposing medical imagery onto patients' anatomy. Similar to this, augmented reality (AR) in education makes learning more engaging by giving students immersive experiences that go beyond standard textbooks. These experiences include interactive simulations and 3D representations that help students better comprehend and remember difficult topics [7].

II.RELATED WORKS

The act of identifying a target in a picture is known as image recognition technology, which involves simulating human senses with computer technology to finish the image recognition and comprehension process [7]. Target object detection in pictures is a major area of study for image recognition and is extensively applied in the Internet, transportation, and security domains [8].

Object detection and object recognition are two subcategories of object recognition [9]. In contrast to object detection, which requires both the precise position of the object and a feature description of the object in the picture, object recognition just requires a description of the target object's features in the image.

As a result, object identification necessitates examination of the object structure in addition to describing the target.

Feature learning is the primary focus of object recognition. A visual word packet model based on picture recognition was developed by pertinent researchers, who also used the Bag of Words (BoW) model from text recognition to the field of image object identification. It has greater robustness and discrimination thanks to the underlying feature extraction and feature coding.

Next, the feature set aggregation process yields the feature expression for the entire picture. Finally, picture recognition is accomplished by the support vector machine.

The word packet model-based image recognition algorithm extracts the image's underlying features by performing edge and corner detection on the image's texture and edge features, respectively, using the H augmented reality ris-Laplace operator and the Placian Gaussian operator. For feature description, in addition to the Scale Invariant Feature Transform (SIFT) technique, the local feature descriptor The properties of the two-dimensional coordinate distribution histogram around the feature points may also be counted using Spin Image.

In order to extract features at various sizes, the algorithm model additionally uses a dense feature extraction technique based on a fixed grid.

The MNIST and CIFAR-10 databases are used in this article's experiments, which examine convolution kernel size and number, pooling size and technique selection, parameter update algorithm selection, activation function selection, and data augmentation. The findings are analyzed. Specifically, the following is a summary of the technical contributions made by this article:

First, we integrate the XGBoost method, a well-known machine learning technique, with the multi-layer characteristics of the convolutional neural network model in deep learning.

The convolutional neural network's extracted features are serially fused, and the dimension is reduced using Principal Components Analysis (PCA) technology.

Second: In the dynamic image recognition experiment on these two datasets, 64 chosen cores, a 33-by-33 receptive field, a 33-by-33 random pool, and the application of stochastic gradient descent (SGD) with momentum are all employed. To get a decent recognition result, the optimization technique employs a 5-layer deep convolutional neural network, increases the amount of input, and utilizes Relu as the activation function.

III. METHODOLGOIES

The image that the user sees in the augmented reality system adjusts with the viewing angle. The user's position, line of sight, and other data must be precisely tracked by the augmented reality system in real time. The augmented reality system's functionality is determined on how well the tracking technology performs.

Four primary types of sensor-based tracking technologies are frequently utilized in augmented reality systems: inertial, magnetic field, optical, and audio tracking technologies.

Typically, a magnetic field tracking system consists of transmitters, receivers, and control components. Electromagnetic induction coils that are mutually orthogonal make up the signal transmitter and receiver. An electromagnetic induction coil is used by the signal generator to create a magnetic field, which is then received by the signal receiver, which creates a matching induced current. Based on the receiver's current signal, the control unit's algorithm may determine the tracking target's position and orientation in relation to it.

The magnetic field tracking technique has greater system refresh rates and improved real-time performance. It is also not restricted by sight lines or obstructions and is not affected by other items other than conductive materials and magnets. Furthermore, customers will find the magnetic field tracking system's compact and light sensing device useful. Magnetic field tracking technique is mostly used in small-scale augmented reality applications that do not require conductive magnets.

There are several types of light sources and photosensitive devices used in optical tracking technology. The transmission medium of the optical tracking system is an optical signal, therefore the signal reception speed is rapid and the refresh rate is high, which is appropriate for situations with high real-time needs. The photosensitive device can be a regular camera or a photodiode. Nonetheless, the absence of any obstructions between the sensor and the optical element is necessary for the optical tracking system to function. Furthermore, optical tracking technologies come at a somewhat high cost.

Integrating virtual data with real-world settings is a fundamental challenge in the design of augmented reality systems. The user is presented with the augmented reality system's final result via a variety of channels, and this display effect dictates the user's intuitive perception. As a result, display technology plays a critical role in augmented reality applications.

The user's line of sight is obscured by the helmet, as seen in the video view, leaving one or two cameras in place to capture the actual situation. The scene synthesizer creates a fusion result from the camera footage and graphics, which is then finally sent to the display in front of the user. The user may see both the virtual picture that the synthesizer has projected and the actual scene in front of them via it.

IV. CNN-XGBoost AUGMENTED REALITY DYNAMIC IMAGE RECOGNITION MODEL

The extraction and expression of features are crucial for target identification tasks, and the accuracy of the feature expression data derived from the data has a significant impact on task completion.

Manual extraction is the basis for traditional data extraction features.

Manual extraction techniques can also be used for applications involving tiny data sets and feature sizes. However, manual feature extraction is likely to result in missing feature information or erroneous feature extraction when the volume of data and the dimensions of the data features are significant. It frequently takes practice and a lot of trials to progressively find an appropriate way to extract characteristics using manual extraction. The feature dimensions of picture data are quite vast due to the introduction of enormous data sets. The challenge of feature extraction is resolved by the concept and advancement of deep learning. Deep learning can identify complicated structures in data and has high feature learning skills.

Convolutional neural networks extract features from images quite well. Complex and abstract advanced characteristics of image data may be recovered by the network model layer extraction, which can then be utilized to extract the vital information of the image data and be used to identify various targets within the image. Without any kind of processing, the image is sent directly into the convolutional neural network. In order to accomplish the goal of feature extraction from the image data, the convolutional neural network directly learns from the picture data and gains knowledge about its features.

The CIFAR-10 picture is sent into the distillation learning model of Alex Net, which is used to extract image characteristics. A sequence of procedures, including pooling and convolution layers, may be used to retrieve the characteristics of various layers.

Only the characteristics from the network model's final layer are utilized to train the Alex Net model. These features are then sent into the softmax layer, which is identified by the softmax recognizer. It's possible that the feature information is incomplete.

Multilayer feature information is employed for identification in order to improve the model's performance and capacity for generalization. We can determine that the convolutional layer and the pooling layer are mostly responsible for extracting the features from the picture data by analyzing the structure of the convolutional neural network. The retrieved characteristics are transformed into a one-dimensional output feature vector—the data sample—by the last fully connected layer. The characteristics that the convolutional neural network's later layers extract are often more pertinent to the original input and more suited to be features that differentiate data. As a result, the features in the last three layers of the Alex Net model are fused, and the resulting fused features may be described using the formula below. The fused feature vector is easily over-trained and has large dimensions. The most representative features are chosen to avoid over-fitting, and PCA technology is utilized for feature dimensionality reduction in order to address the issue of fused feature dimensions being high and feature information being redundant. Fitting decreases computation time and increases calculation speed while improving the model's capacity for generalization.

Known as one of the most significant outcomes of linear algebra, PCA is a technique for reducing the dimensionality of data. It is a straightforward non-parametric technique for taking pertinent information out of complicated data. PCA is a method for reducing large data sets to smaller dimensions and identifying the most crucial features with little further work. Assuming that the feature vector's dimension is two, the original data consists of two features. The covariance matrix of the data set is computed using the PCA method, which also computes the covariance matrix's eigenvalue and eigenvector. The secondary linear component is the feature vector that corresponds to the tiny eigenvalue, and the feature dimension can be decreased to 1.

Converting data from the old coordinate system to the new coordinate system is the fundamental notion behind PCA. The data itself establishes the new coordinate system; that is, the n-dimensional feature is translated to the k-dimensional feature. A novel orthogonal feature, referred to as the major component, is the k-dimensional feature. The direction with the biggest data difference—that is, the highest variance—is used to choose the first principal

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 06 | June - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

component. Choosing the direction with the second highest data difference, the second principle component is orthogonal to the first principal component, and so on.

V. EXPERIMENT AND ANALYSIS

Theoretically, the number of kernels (filters) in the convolutional layer corresponds to the number of feature maps. The network representation feature space grows as the number of kernels increases and more feature maps are extracted. The accuracy of the identification increases with ability strength.

We created a variety of structural models based on CNN-XGBOOST to investigate the impact of core count on convolutional neural network performance. We update its three-layer convolution structure to be: 8-8-32, 16-16-32, 32-32-32, 32-32-64, 64-64-64, 64-64-128 while maintaining its hierarchical structure and other elements. Two datasets were used for the experiments. We changed the size of the three-layer receptive field in accordance with the CNN-XGBOOST structural model in order to investigate the impact of the receptive field's (convolution kernel) size on the convolutional neural network's performance: 10~10, 9~9, 8~8, 7~7, 5~5, 3~3, 2~2. We conduct our experiment on two databases while holding other variables constant.

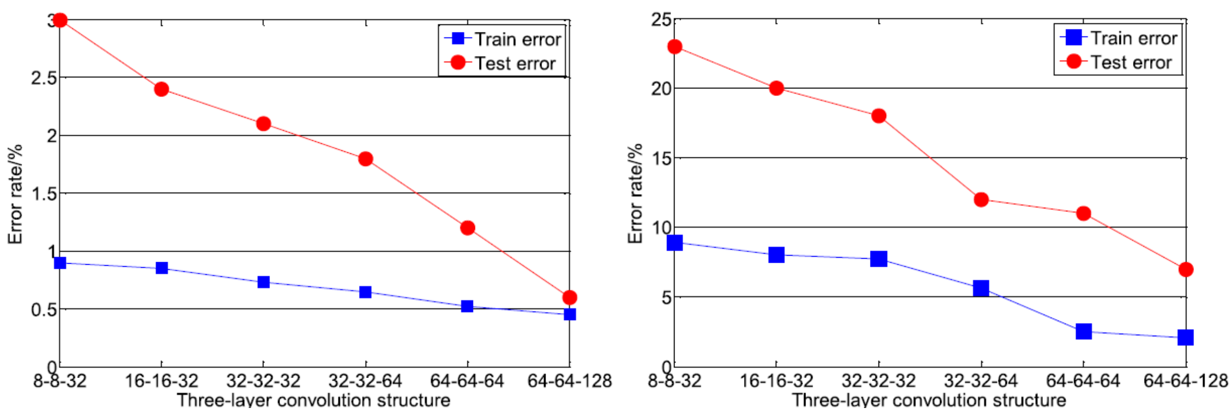

Figure 1. Recognition results of different core number designs.

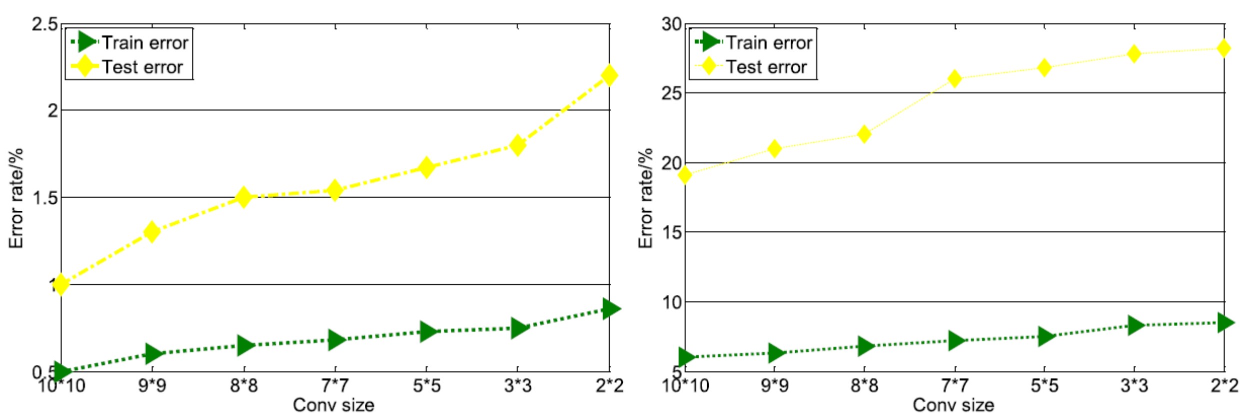

Figure 2. Recognition results of different receptive field sizes.

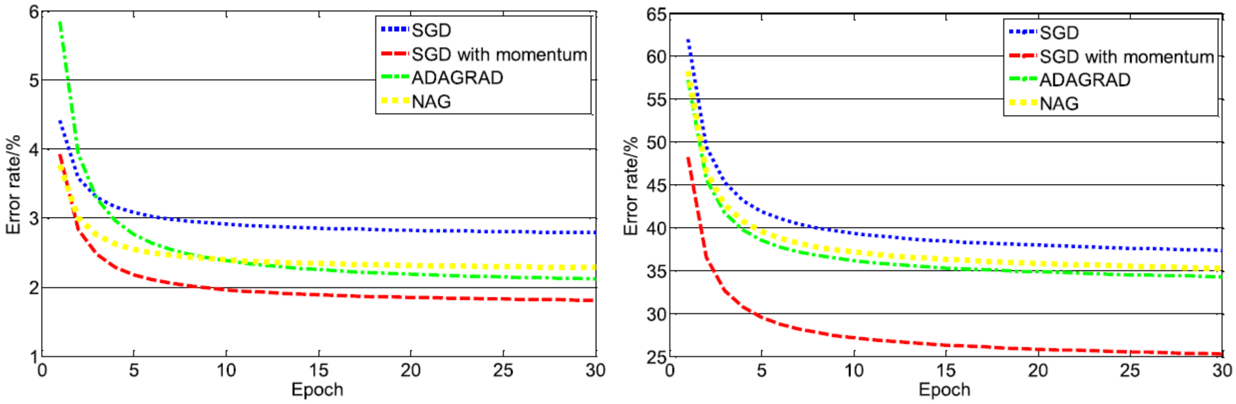

Figure 3. Experimental results of different optimization algorithms.

A gradient technique must be used to update the network parameters in order to solve our neural network and maximize the network cost function. The most popular ones include Nesterov's Accelerated Gradient (NAG), Adaptive Gradient (ADAGRAD), SGD, and SGD with momentum.

In order to evaluate the effectiveness of these techniques on convolutional neural networks, we modify the CNN-XGBOOST model-based parameter update approach by setting the model's learning rate to 0.012, mini-batch to 9, and regularization coefficient to 0.0008.

The test errors for 30 iterations using these optimization techniques on the MNIST and CIFAR-10 databases are displayed in Figures 9(a) and 9(b), respectively.

The experiment shows that while SGD without momentum can gradually reduce the error, the decline speed and effect are general; SGD and NAG with momentum both have better effects than SGD, but SGD with momentum outperforms NAG in the early stages of the gradient decline rate. Because ADAGRAD's falling speed is so steady, the curve is smoother and the effect is superior.

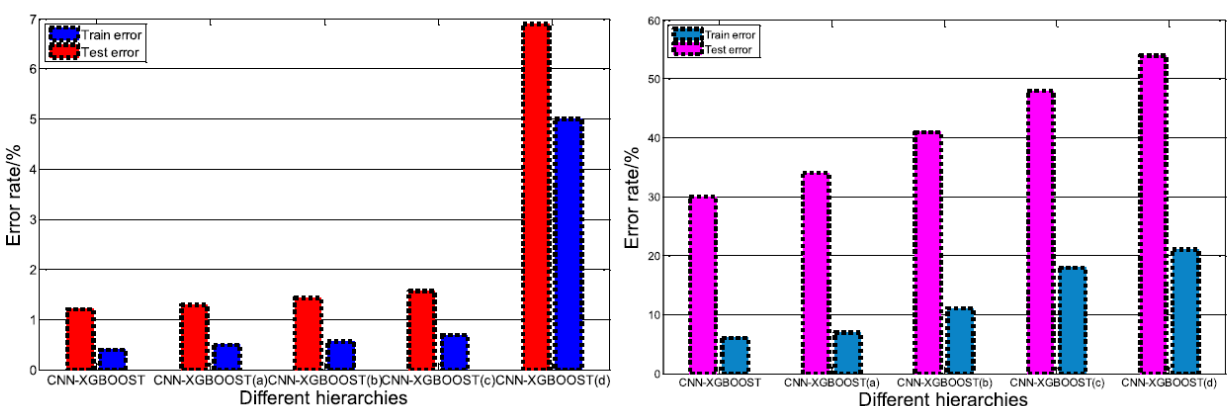

Figure 4. Identification results of different hierarchical structures.

By employing the aforementioned five architectures, we conduct recognition tests in the MNIST database while maintaining other parameters same to CNN-XGBOOST. The result of recognition is displayed in Figure 1. On the

CIFAR-10 data set, we conduct recognition tests using the aforementioned five architectures. The result of recognition is displayed in Figure 2.

According to the experimental findings, removing the top fully connected layer has no effect on the outcomes, whereas removing the convolutional pooling layer causes a notable drop in performance. This demonstrates how the convolutional pooling layer affects the outcome. When the two-layer convolution pooling layer is removed, the error rises especially quickly.

Experiments also confirmed that a network's depth greatly influences its performance. The accuracy increases with network depth, and distinct learning characteristics result from deeper networks. But the greater depth a network has, the more parameters it has, and the more complex it becomes. Based on its data collection, the selection of rising levels is likewise optimized.

VI. CONCLUSION

The goal of augmented reality technology, which combines virtual and physical technologies, is to precisely record computer-generated virtual information into real-time scene photographs that are collected in real-time to create an improved image that can be displayed to consumers, improving their sensory experience. In order to address the issue of insufficient feature information resulting from the standard neural network's use of only the last layer of features for recognition, this article integrates the convolutional neural network with the XGBOOST algorithm in integrated learning. In order to maintain more feature information for recognition, this article employs the technique of fusing multi-layer features. The XGBOOST recognizer, which has strong recognition effect, is then used for recognition. For databases such as MNIST and CIFAR-10, random pooling is the preferred pooling approach.One interesting direction for educational innovation is the creation of AR-based image recognition teaching models. Teachers may design immersive and interactive learning environments that meet the requirements and preferences of a wide range of students by utilizing AR's immersive and interactive features. Despite certain obstacles, further investigation and funding for research and development are necessary to fully realize the educational potential of augmented reality.

REFERENCES

[1] F. Cheng, ``Image recognition technology based on deep learning," Wire-less Pers. Commun., vol. 102, no. 2, pp. 1917-1933, Jan. 2018.

[2] X. Wang, W. Zhang, X. Wu, L. Xiao, Y. Qian, and Z. Fang, ``Realtime vehicle type classification with deep convolutional neural networks," J. Real-Time Image Process., vol. 16, no. 1, pp. 5-14, Aug. 2019.

[3] P. R. Jeyaraj and E. R. S. Nadar, ``Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," J. Cancer Res. Clin. Oncol., vol. 145, no. 4, pp. 829-837, Jan. 2019.

[4] S. Law, C. I. Seresinhe, Y. Shen, and M. Gutierrez-Roig, ``Street-frontagenet:Urban image classification using deep convolutional neural networks,"Int. J. Geographical Inf. Sci., vol. 34, no. 4, pp. 681-707, Dec. 2020.

[5] Ravindra Changala, “Secured Activity Based Authentication System” in " in Journal of innovations in computer science and engineering (JICSE), Volume 6, Issue 1,Pages 1-4, September 2016.ISSN: 2455-3506.

[6] Ravindra Changala, “Automated Health Care Management System Using Big Data Technology”, at Journal of Network Communications and Emerging Technologies (JNCET), Volume 6, Issue 4, April (2016), 2016, pp.37-40,ISSN: 2395-5317, ©EverScience Publications.

[7] Ravindra Changala, "Retrieval of Valid Information from Clustered and Distributed Databases" in Journal of innovations in computer science and engineering (JICSE), Volume 6, Issue 1, Pages 21-25, September 2016.ISSN: 2455-3506.

[8] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, ``Cascaded recurrent neural networks for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 8, pp. 5384-5394, Aug. 2019.

[9] Y. Qin, L. Bruzzone, B. Li, and Y. Ye, ``Cross-domain collaborative learning via cluster canonical correlation analysis and random walker for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 6, pp. 3952-3966, Jun. 2019.

[10] X. Zhu, Z. Li, X.-Y. Zhang, P. Li, Z. Xue, and L. Wang, ``Deep convolutional representations and kernel extreme learning machines for image classication," Multimedia Tools Appl., vol. 78, no. 20, pp. 29271-29290, Nov. 2019.

[11] X. He and Y. Chen, ``Optimized input for CNN-based hyperspectral image classification using spatial transformer network," IEEE Geosci. Remote Sens. Lett., vol. 16, no. 12, pp. 1884-1888, Dec. 2019.

[12] J. W. Kim and P. K. Rhee, ``Image recognition based on adaptive deep learning," J. Inst. Internet, Broadcast. Commun., vol. 18, no. 1, pp. 113-117, Feb. 2018.

[13] Ravindra Changala, Framework for Virtualized Network Functions (VNFs) in Cloud of Things Based on Network Traffic Services, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169 Volume 11, Issue 11s, August 2023.

[14] Ravindra Changala, Optimization of Irrigation and Herbicides Using Artificial Intelligence in Agriculture, International Journal of Intelligent Systems and Applications in Engineering, Volume 11, Issue 3), July 2023.

[15] Ravindra Changala ,Development of Predictive Model for Medical Domains to Predict Chronic Diseases (Diabetes) Using Machine Learning Algorithms And Classification Techniques, ARPN Journal of Engineering and Applied Sciences, Volume 14, Issue 6, 2019.

[16] S. Mahdizadehaghdam, A. Panahi, H. Krim, and L. Dai, ``Deep dictionary learning: A PARametric NETwork approach," IEEE Trans. Image Pro- cess., vol. 28, no. 10, pp. 4790-4802, Oct. 2019.

[17] S. Zhou, Z. Xue, and P. Du, ``Semisupervised stacked autoencoder with cotraining for hyperspectral image classication," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 6, pp. 3813-3826, Jun. 2019.

[18] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, ``Deep learning for hyperspectral image classication: An overview," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 9, pp. 66906709, Sep. 2019.

[19] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, ``Visual attentiondriven hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 10, pp. 8065-8080, Oct. 2019.

[20] S. Antholzer, M. Haltmeier, and J. Schwab, ``Deep learning for photoacoustic tomography from sparse data," Inverse Problems Sci. Eng., vol. 27, no. 7, pp. 987-1005, Sep. 2019.

[21] Y. Niu, Z. Lu, J.-R. Wen, T. Xiang, and S.-F. Chang, ``Multi-modal multiscale deep learning for large-scale image annotation," IEEE Trans. Image Process., vol. 28, no. 4, pp. 1720-1731, Apr. 2019.

[22] Ravindra Changala, "Evaluation and Analysis of Discovered Patterns Using Pattern Classification Methods in Text Mining" in ARPN Journal of Engineering and Applied Sciences, Volume 13, Issue 11, Pages 3706-3717 with ISSN:1819-6608 in June 2018.

[23] Ravindra Changala "Pattern Deploying Methods for Text Mining" in International Journal of Soft Computing, Volume 13, Issue 2, pages 61-68 with ISSN: 1816-9503 in June 2018.

[24] Ravindra Changala "A Survey on Development of Pattern Evolving Model for Discovery of Patterns in Text Mining Using Data Mining Techniques" in Journal of Theoretical and Applied Information Technology, August 2017. Vol.95. No.16, ISSN: 1817-3195, pp.3974-3987.

[25] K. Z. Haider, K. R. Malik, S. Khalid, T. Nawaz, and S. Jabbar, ``Deepgender: Real-time gender classification using deep learning for smartphones," J. Real-Time Image Process., vol. 16, no. 1, pp. 15-29, Sep. 2019.

[26] Ravindra Changala, "Optimizing 6G Network Slicing with the EvoNetSlice Model for Dynamic Resource Allocation and Real-Time QoS Management", International Research Journal of Multidisciplinary Technovation, Vol 6 Iss 4 Year 2024, 6(4) (2024) 325-340.

[27] Ravindra Changala, "Real-time Anomaly Detection in 5G Networks through Edge Computing", 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS),|979-8-3503-6118-6/24/©2024IEEE|DOI: 10.1109/INCOS59338.2024.10527501.

[28] Ravindra Changala, "Enhancing Quantum Machine Learning Algorithms for Optimized Financial Portfolio Management", 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 979-8-3503-6118-6/24/$31.00 ©2024 IEEE.

[29] Ravindra Changala, "Biometric-Based Access Control Systems with Robust Facial Recognition in IoT Environments", 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), | 979-8-.3503-6118-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/INCOS59338.2024.10527499.

[30] Ravindra Changala, "Deep Learning Techniques to Analysis Facial Expression and Gender Detection", IEEE International Conference on New Frontiers In Communication, Automation, Management and Security(ICCMA-2023), 979-8-3503-1706-0/23©2023IEEE|DOI: 10.1109/ICCAMS60113.2023.10525942.

[31] Liu, L., Li, Y., Peng, S., & Fu, S. (2018, June). Repetition Counting for Exercise Recognition with Deep Learning. In 2018 IEEE Conference on Computer Vision and Pattrn Recognition (CVPR) Workshops (pp. 1773-1782).