AIMS Mathematics, 8(9): 21554–21571.
DOI: 10.3934/math.20231098
Received: 10 May 2023
Revised: 18 June 2023
Accepted: 26 June 2023
Published: 07 July 2023

http://www.aimspress.com/journal/Math

Research article

Vision graph neural network-based neonatal identification to avoid swapping and abduction

Madhusundar Nelson[1], Surendran Rajendran[1,*] and Youseef Alotaibi[2]

[1] Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, 602105, India
[2] Department of Computer Science, College of Computer and Information Systems, Umm Al-Qura University, Makkah, 21955, Saudi Arabia

* **Correspondence:** Email: surendran.phd.it@gmail.com; Tel: +919444013042.

Abstract: Infant abductions from medical facilities such as neonatal switching, in which babies are given to the incorrect mother while in the hospital, are extremely uncommon. A prominent question is what we can do to safeguard newborns. A brand-new vision graph neural network (ViG) architecture was specifically created to handle this problem. Images were divided into several patches, which were then linked to create a graph by connecting their nearest neighbours to create a ViG model, which converts and communicates information between all nodes based on the graph representation of the newborn's photos taken at delivery. ViG successfully captures both local and global spatial relationships by utilizing the isotropic and pyramid structures within a vision graph neural network, providing both precise and effective identification of neonates. The ViG architecture implementation has the ability to improve the security and safety of healthcare facilities and the well-being of newborns. We compared the accuracy, recall, and precision, F1-Score, Specificity with CNN, GNN and Vision GNN of the network. In that comparison, the network has a Vision GNN accuracy of 92.65%, precision of 92.80%, F1 score of 92.27%, recall value of 92.25%, and specificity of 98.59%. The effectiveness of the ViG architecture was demonstrated using computer vision and deep learning algorithms to identify the neonatal and to avoid baby swapping and abduction.

Keywords: computer vision; deep learning; vision graph neural network; isotropic pyramid; swapping
Mathematics Subject Classification: 68Q32, 68T40, 68T07, 92D30

1. Introduction

Babies who are mistakenly or intentionally swapped with each other at birth or very shortly after known as newborns switched at birth. As a result, the babies are advertently raised by parents who are not their real parents. It is terrible to switch babies and is difficult to picture returning home with a baby other than your own after multiple months of dreaming of hugging your child. The fact that there is no practical method to predict how many babies would potentially be swapped a year is crucial to recognize. Some hospitals provide cutting-edge methods to prevent your child from being switched to allay that concern. To prevent neonatal swapping, hospitals employ a variety of techniques, including foot-printing, banding, tags that beep, identifying uniforms and nametags, and more [1].

Living together prevents your infant from leaving your side without your consent and guarantees that you are always aware of is the location of your child; if you need to travel, you are free to take the newborn with you. If you are unable to accompany your child, you can entrust the care of your child with someone else (i.e., your partner). This ensures that your child is always in your sight and gives you the chance to influence any tests, treatment, or seemingly unimportant aspects of your child's hospital stay [2]. You can ask to see the baby's name tag both before and after they leave, just in case they become separated for any reason. Additionally, you should request to examine the official ID tag of whomever interacts with your baby [3]. Moreover, you can take a mental note of the appearance of your infant; this may be useful in locating your infant. Although it isn't always possible, it's important to consider things like the amount of hair, birthmarks, etc [4].

Compared to adult datasets, datasets specifically focused on neonatal images are relatively limited. Collecting a comprehensive dataset of neonatal images is more challenging due to ethical and privacy considerations, as well as the difficulty of obtaining consent from parents or guardians. Neonates have distinct physical characteristics that differ from those of adults. Their facial features, skin texture, and other physical attributes are significantly different, making it challenging to directly apply existing adult identification technologies. Neonatal faces are often underdeveloped, and their appearance can rapidly change as they grow. Their facial features, including size, shape, and proportions, considerably change during the first few months of life. This variability makes it more difficult to establish consistent and reliable identification models compared to adults, whose facial features are relatively stable over time.

Infant abductions are any kidnappings involving children under the age of one. This kind of kidnapping can take many different forms, such as a stranger stealing the child from the hospital or a non-custodial parent taking the youngster. Within two hours of birth, hospitals should capture the baby's footprint, obtain a color picture of them and document their physical evaluation. Staff should be required to wear updated visible color-photographed ID badges [5]. Additionally, staff who work directly with infants should be required to wear a second form of distinctive ID, such as a badge with a pink background, as shown in Figure 1.

Figure 1. Neonatal labor room.

Concerning modern computer vision systems, convolutional neural networks (CNNs) were formerly the most effective industry concerning productivity. Recently, transformers were introduced for visual activities with a competing attention mechanism [6]. Without convolution or self-attention, Multilayer perceptron (MLP)-based vision models are also capable of performing well. These developments eleviate vision models to previously unheard-of levels. The input image is handled differently by various networks. The image data is often shown as a systematic Euclidean network of pixels. CNNs introduce shift invariance and locality while applying a sliding window to the image [7]. Recent vision transformers, such as MLP, treat images as a series of patches. For instance, ViT creates a sequence with a length of 196 using 16 × 16 patches to break up a 224 × 224 image into smaller blocks, as shown in Figure 2.

Figure 2. Grid layout.

The image has been processed in a more adaptable manner than the standard grid or sequence representation. Object recognition in images is a fundamental function of computer vision [8]. Older networks such as ResNet and ViT, which frequently use either a grid or sequence design, are stiff and redundant in their processing of these items because their shapes are frequently asymmetrical and non-quadrate. We can produce a diagram showing the image's representation as a grid, sequence, and graph [9]. Only the spatial position determines the order of the pixels or patches in the grid layout. The order structure converts the 2D image into a series of patches. In the network structure, the bulges are connected by their content rather than their local position [10]. Grids and sequences are examples of particular cases of the generic data structure known as a graph. Visual perception is more adaptable and successful when an image is viewed as a graph, as shown in Figure 3.

Figure 3. Sequence layout.

The vision graph neural network for imaged goals is constructed from the graph illustration of pictures. When considering each pixel as a node, which would result in an excessive number of nodes (>10K), the input image is broken into several patches, with each patch being treated as a node. After creating the network of image patches, the ViG model is applied to all nodes to transform and exchange the data. The ViG should go into great detail on the procedures and methods utilized to create the ViG architecture that is suggested for neonatal identification. The methodology section should contain the following important information in terms of making a graph.

For nodes, each newborn image in the ViG architecture is represented as a node in the graph. These nodes represent the visual information that convolutional layers used to extract from the newborn photographs.

For edges, the graph's edges show the connections between neonatal nodes. Nodes with similar visual characteristics or spatial interactions are connected by using proximity or similarity measurements to construct links between them.

In terms of training the graph neural network, the ViG design incorporates graph convolutional layers during training to maintain the graph structure. These layers compile data from nearby nodes, allowing the model to gain knowledge from the linked neonatal representations. Through the use of graph convolutions, the model gains the ability to incorporate spatial dependencies and discriminative characteristics.

By using the connections and weights it has learnt during the training phase, the ViG architecture preserves the graph's structure during inference. The model uses the conserved associations in the graph convolution processing of the input newborn photos to deduce the identities of the neonates.

In computer vision, the term "isotropic" refers to a quality or feature of an image or feature that maintains its integrity when subjected to transformations like rotation, translation, and scaling. It signifies that regardless of the orientation or scale, an image's or a feature's features or characteristics remain the same. Given that it enables reliable and invariant feature extraction, this trait is frequently desired in computer vision tasks.

When referring to a multi-scale representation of an image or feature hierarchy, the term "pyramid structures" is usually used. In a pyramid structure, the original image or feature is represented at several scales, with each level capturing varying degrees of detail. The idea that things or features can be found or recognized at various scales is where the term originates. In tasks such as object identification, picture segmentation, and image recognition, a pyramid structure is frequently utilized because it enables the extraction of information at various resolutions.

Graph and FFN (feed-forward network) modules make up the two main components of ViG's basic cell. The graph module is built using graphical convolution to process graph information, as shown in Figure 4.

In order to mitigate the over-smoothing issue brought on by traditional GNNs, an FFN module is utilized to change the node features and to encourage node diversity. We can construct both isotropic and pyramidal ViG models using the graph and FFN modules. Results showed that the ViG model performs well on visual tasks like object and image detection. For instance, on the ImageNet classification task, Pyramid, ViG-S exceeds the ideal CNN (ResNet), MLP (CycleMLP), and transformer (Swin-T) with comparable FLOPs, achieving the highest accuracy. To the best of our knowledge, this research represents the successful application of graph neural networks to significant visual tasks [11,12]. The use of metaheuristic algorithms has expanded in recent years as various problems have become more complex. In the past, scholars dealt with the deterministic and local challenging-to-trap optimization problems using mathematical techniques. Practical optimization

problems, including text processing, community detection, feature selection, optimization problems, setting machine learning parameters, etc., almost always entail inherent complexity limits and a number of design considerations, including nonlinearity and convexity.

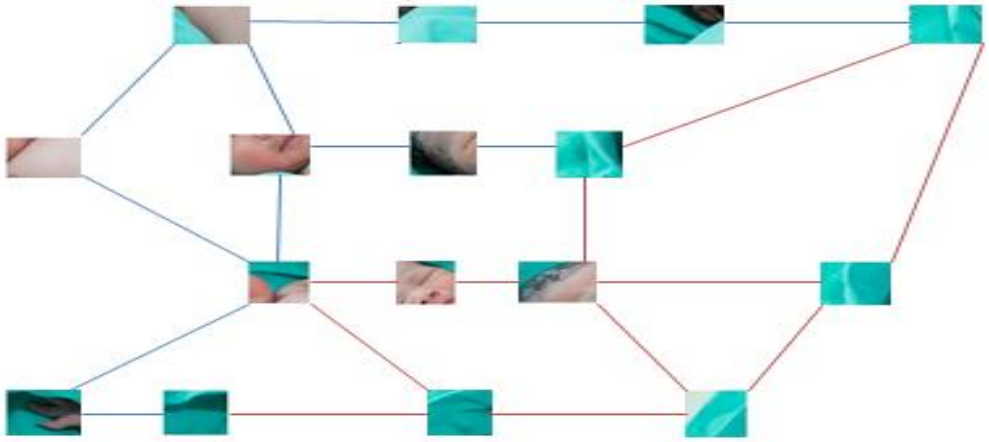

Figure 4. Graph layout.

To create a novel ViG architecture that would mine graph-level characteristics for use in visual applications and represent the image as a graph structure, the network builds its isotropic and pyramid structures utilizing a range of model sizes. The network divides the image into numerous patches, which are then joined to form a graph by connecting their nearest neighbours, thereby building our ViG model, which is based on the graphical representation of the newborn's photographs acquired during birth, to convert and exchange information among all nodes. Extensive trials on computer vision methods and deep learning algorithms were carried out to detect newborns and object identification to prevent infant swapping and abduction, demonstrating the effectiveness of our ViG architecture.

2. Related work

First, we re-examine the computer vision's foundational networks. Next, we examine the evolution of graph neural networks, particularly GCN, and its use in visual tasks. In the past, convolutional networks remained the ordinary network architecture for computer vision. Since LeNet, CNNs efficiently consumed to a variety of pictorial submissions, counting semantic subdivision, entity identification, and image classification. Over the past decade, CNN architecture has made remarkable progress. ResNet, MobileNet, and NAS are a few examples of the representative efforts. For visual activities, a vision transformer was introduced in 2020. The performance of visual tasks was subsequently recommended to be improved by using different versions of ViT. The main innovations include pyramid architecture, local attentiveness, and location encoding. Vision transformers served as an inspiration for MLP, which is also being studied in terms of computer vision. In general, MLP may work hard for visual tasks like object detection and segmentation, and competes at a high level with appropriately designed modules.

The three primary computer vision applications of GCN are generating scene graphs, classifying point clouds, and identifying activities. A point cloud is a collection of three-dimensional (3D) points in space, generally acquired by LiDAR scans. Using GCN, point clouds have been categorized and

segmented. Scene graph generation typically combines an object detector and GCN to assess the input image and create a graph with the items and their relationships. GCN was used for the human action recognition problem by analyzing the naturally produced graph of connected human joints. Only naturally constructed graphs are capable of handling specific visual tasks by GCN. The picture facts must be supported by a GCN-based backing network technique for a general display in computer vision.

3. Proposed work

For neonatal intensive care units (NICUs), research is now being performed on the development of non-contact patient monitoring software that primarily makes use of face image processing. Recent research has used information from newborn photographs to identify individuals by their faces in order to prevent neonatal swapping and kidnapping. For these applications, the region of interest (ROI) must be correctly identified as the infant's face. Until a face is discovered, the image is rotated using the dlib and OpenCV visual libraries. Cropping is a technique in image preprocessing that is used to consistently place the face by using facial landmarks such the eyes, brows, nose, mouth, and jawline. The hardware and the software utilized includes an Intel (R) Core (TM) i7-8750 processor and an NVIDIA GeForce GTX 1080 GPU. Python 3.8.13 and Keras 2.10.0 were the two programming languages employed.

3.1. Data augmentation

A group of methods was utilized for theoretically increasing the number of data points by generating additional ones from the current ones. Deep learning algorithms can either produce new data or subtly alter currently existing data. By adding new and intriguing examples to training datasets, it enhances the deep learning models' functionality and output. When the dataset is extensive and huge enough, a deep learning model can perform more precisely and successfully. Processing operations for data augmentation include padding, random erasing, vertical and horizontal flipping, translation, cropping, zooming, rotating, rescaling, darkening and lightening, grey scalability, changing contrast, and cropping. Examples of image processing techniques include brightness adjustments, the addition of Gaussian noise, rotations between −10 and 10 degrees, scaling between 0.5 and 1.5 times, and channel switching. Data preparation is the process of converting unprocessed data into a format that can be used and understood. Real-world or raw data frequently contains errors, omissions, and inconsistent formatting. The precision and efficacy of analytical methods for datasets are improved for data preparation by addressing these issues.

The detection of HSV colors in human skin was compared to the YCbCr color space. A human's skin color can be determined by distinguishing skin from non-skin pixels. The HSV color space characteristics for skin color identification is an important tool for differentiating between skin and non-skin points in a photograph. YCbCr skin recognition often uses neural network technology. The analysis's findings demonstrate that the performance of the YCbCr color space model is superior to that of the RGB color space model. The outcome of blurring an image with a Gaussian function is a Gaussian blur, also known as Gaussian smoothing. It is common practice to remove noise and to reduce detail from images using the Gaussian Filter with Python and OpenCV.

GNN structure for visual representation learning is used to incorporate vision into a graph from an image. Facet points have a complicated topological distribution, making grid or sequence structures

inadequate to represent them. On the other hand, because they are formed of nodes and edges, graphs can handle complex unstructured data. The relationship between each feature point and each node can be calculated using the distance between each node in the network [13–15]. Any binary relational system that can be visualized as a graph is used to express the relationship between objects. A graph's vertex and edge elements can be used to represent things and the connections between them [16–18]. This strategy develops representations of hidden layers that encode the topology of the local network and node properties such as the number of graph edges rises, as shown in Figure 5.

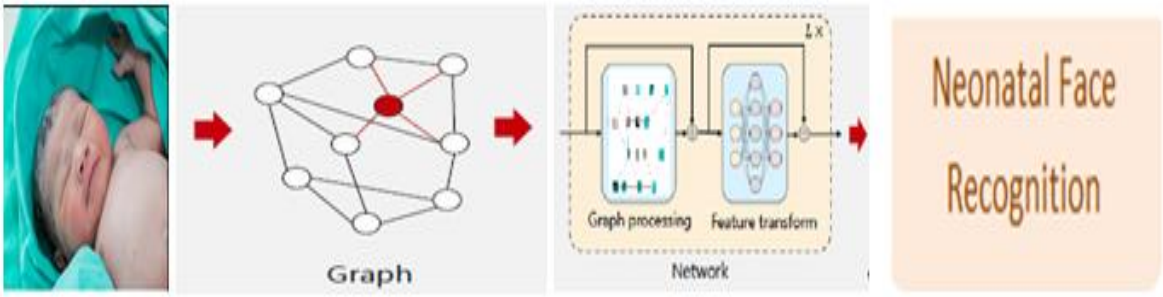

Figure 5. The framework of the ViGNN model.

3.2. ViG block

The proposed ViG architecture for neonatal identification is represented as a node in the graph, making a graph node represents each newborn image in the ViG architecture. These nodes represent the visual information that the convolutional layers use to extract from the newborn photographs. The graph's edges show the connections between neonatal nodes. Nodes with similar visual characteristics or spatial interactions are connected by using proximity or similarity measurements to construct links between them for the graph neural network in Training; the ViG design incorporates graph convolutional layers during training to maintain the graphical structure. These layers compile data from nearby nodes, allowing the model to gain knowledge from the linked neonatal representations. Through the use of graph convolutions, the model gains the ability to incorporate spatial dependencies and discriminative characteristics.

ViG makes sure that both local and global spatial information is efficiently collected by incorporating isotropic and pyramid structures, allowing the extraction of fine-grained visual features. To extract multi-scale characteristics from newborn pictures, the isotropic structure in ViG employs a number of convolutional layers. This structure gives the model the ability to gather background knowledge and spot minute characteristics that are essential for precise identification. On the other hand, the pyramid structure makes use of layer pooling and skip connections to capture hierarchical representations at various scales. It is now possible to differentiate and fully understand neonates. The graph-based approach of ViG is particularly beneficial for the task of neonatal identification due to several reasons such as capturing complex relationships, handling unique challenges in neonatal imaging, considering contextual understanding, and offering flexibility and scalability contribute to its superior performance.

First, we divide an image of size H x W x 3 into N patches to obtain Y = [Y1, Y2,..., YN], where D is the feature measurement and i = 1, 2,..., N. By converting each patch into a feature vector Yi € RD, these features can be thought of as a collection of unarranged nodes, given by the notation V = {v1, v2, ….,vN}., … Next, we find the K closest neighbors of each node vi, N(vi), then improve an edge eji

fixed from vj to vi for all vj, € N(vi). The resulting graph is G = (V, E), where E stands for all edges. The process of creating a graph is indicated as G = G(Y) in the following paragraphs. Next, we investigate how to use GNN to extract information from the image by viewing it as a graph of the data [19,20].

Graphing the image has the following advantages: 1) a generalized data structure known as a graph, the grid and the arrangement can be thought of as a particular instance of the graph; 2) a graph is more flexible than a grid or order to represent a complicated object since an entity in an image often has an irregular shape and is not quadrate; 3) more flexible than a grid or list, a graphical structure can connect an object's sections, which can be thought of as an arrangement of parts (for example, a human can be roughly divided into the head, upper body, arms, and legs); and 4) modern GNN research [21–23] can be applied to tackle visual challenges.

As a general rule, graph-level processing begins with the feature Y, € RNxD. First, we create a graph based on the following topography: g = G(Y). By pooling features from its neighboring nodes, a graph convolutional layer can communicate facts amongst the nodes. In particular, the graphical convolution functions are as follows:

$$g|=k(G,W)Update(Aggregate(g,Wagg),Wupdate) \tag{1}$$

where Wagg and Update are the operations' learnable weights for aggregating and updating, respectively. In more detail, the aggregation operation groups the characteristics of neighboring nodes to compute the representation of a node, and the apprise action additionally combines the aggregated feature

$$Y|i = h(y\ i, g(yi,N(yi),Wagg),Wupdate), \tag{2}$$

where N(y|i) is the collection of y|i 's neighbor nodes. For its simplicity and effectiveness, in this case, we use the max-relative graph convolution.

$$g(.) = y||i = [yi, max\ (\{yj - yi \mid j \in N(yi\ )g\}], \tag{3}$$

$$h(.) = y|i = y||\ iWupdate. \tag{4}$$

This omits the preference term. The processing at the graphical level described above can be represented as y| = GraphConv (y). The accumulated feature y|| i is divided into h faces, or face1, face2, …, faceh, and is updated with various weights for each face. The final values are the concatenation of all faces, which can be updated concurrently:

$$Y|i = [face1W1update, face2W2update, facehWh\ update]. \tag{5}$$

The model may update data in several representation subspaces for the benefit of feature diversity thanks to multi face update operation. For extracting aggregated features from the graph data, prior GCNs frequently used numerous graph convolution layers repeatedly. Visual recognition performance will suffer due to the over-smoothing phenomenon of deep GCNs, which will make node characteristics less distinguishable, where y = arg minx||Y, 1yT|| and the diversity is evaluated as ||Y - 1y-T||. To solve this issue, our ViG block should have more feature transformations and nonlinear activations [24,25].

We use a linear layer both before and after the graphical convolution to safeguard the node topologies into the same field and increase the feature set. After the graph convolution, a nonlinear activation function is applied to prevent layer collapse. The improved module is known as the grapher module. In reality, the grapher module can be described as the following given input feature Y, € RNxD:

$$X = \sigma(\text{GraphConv}(YWin))Wout + Y. \quad (6)$$

The bias term and the activation function are omitted, such as ReLU and GeLU, if the fully connected layer weights are X, €RND, Win, and Wout. Employed feed-forward networks (FFN) on each node are employed to reduce the over-smoothing phenomenon and further support the feature transformation capability. The FFN module has two entirely connected layers and is a simple multilayer perceptron,

$$Q = \sigma(XW1)W2 + X\ . \quad (7)$$

In this case, the bias term is disregarded, while Q, € RND, W1, and W2 are the weights of layers that are totally linked. Usually, FFN's concealed measurement is higher than D. After every fully connected layer or graphical convolution layer in the Grapher and FFN modules, batch normalization—which is skipped in Equations 6 and 7 for conciseness—is applied. The main building element for creating a network is the ViG block, which is made up of a stack of Grapher and FFN modules.

CNNs are likely to employ a pyramid design, although the generally used transformer in computer vision typically has an isotropic architecture (e.g., ViT) (i.e., ResNet). We construct isotropic and pyramidal network topologies for ViG to thoroughly compare them to other varieties of neural networks. Isotropic architecture refers to a main body that has characteristics like ViT and ResMLP that are the same size and form throughout the network. We create three distinct isotropic ViG architectures, viz., ViG-Ti, S, and B, each with a different model size. The node count is set to N = 196. The number of neighbor nodes K increases linearly from 9 to 18 in these three models as the layer depth increases, thereby widening the receptive field. The default setting for the number of heads in Table 1 is h = 4.

Table 1. Variant of isotropic ViG architecture.

Prototypical	Deepness	Measurement (D)	Params (M)	Flops (B)
ViG-T	12	192	7.2	1.4
ViG-S	16	320	22.9	4.7
ViG-B	16	640	87.1	17.9

The FLOPs are computed for the 224 × 224 resolution picture, where T stands for tiny, S for small, and B for a base. The pyramid design accounts for the multiscale nature of pictures by extracting features like ResNet and PVT that have decreasing spatial sizes as the layer depth increases. According to an empirical study, the pyramid architecture is useful for visual tasks [26,27]. Therefore, utilizing the sophisticated design, we build four distinct pyramid ViG models. Keep in mind that we manage a high number of nodes in the first two stages by using spatial reduction.

In Location encoding, each node feature is given a positional training vector in the command to symbolize the nodes' location information:

$$Yi \quad <\text{—}Yi + ei, \quad (8)$$

where ei € RD. Isotropic and pyramid structures both use the complete positional encoding, as shown in Eq 8. Relative positional encoding is also added for pyramid ViG by using cutting-edge designs like Swing Transformer. The feature distance for nodes i and j will be increased by the qualified positional remoteness amongst them, which is given as eiT, ej, when building the graph in Table 2.

Table 2. Setting details for the Pyramid ViG series.

Level	Output Size	PyramidViG-Ti	PyramidViG-S	PyramidViG-M	PyramidViG-B
Stem	H/4* W/4	ConX3	ConX3	ConX3	ConX3
Level 1	H/4*W/4	$\begin{bmatrix} D=47 \\ E=5 \\ K=9 \end{bmatrix} * 2$	$\begin{bmatrix} D=80 \\ E=4 \\ K=9 \end{bmatrix}$*2	$\begin{bmatrix} D=96 \\ E=4 \\ k=9 \end{bmatrix}$*2	$\begin{bmatrix} D=128 \\ E=4 \\ K=9 \end{bmatrix}$* 2
sample	H/8* W/8	Conv	Conv	Conv	Conv
Level 2	H/8* W/8	$\begin{bmatrix} D=95 \\ E=5 \\ k=9 \end{bmatrix}$* 2	$\begin{bmatrix} D=160 \\ E=4 \\ K=9 \end{bmatrix}$* 2	$\begin{bmatrix} D=192 \\ E=4 \\ K=9 \end{bmatrix}$* 2	$\begin{bmatrix} D=256 \\ E=4 \\ K=9 \end{bmatrix}$* 2
sample	H/16*W/16	Conv	Conv	Conv	Conv
Level 3	H/16* W/16	$\begin{bmatrix} D=239 \\ E=5 \\ K=9 \end{bmatrix}$* 6	$\begin{bmatrix} D=402 \\ E=4 \\ K=9 \end{bmatrix}$*6	$\begin{bmatrix} D=383 \\ E=4 \\ K=9 \end{bmatrix}$*16	$\begin{bmatrix} D=512 \\ E=4 \\ K=9 \end{bmatrix}$*18
sample	H/32* W/32	Conv	Conv	Conv	Conv
Level 4	H/32* W/32	$\begin{bmatrix} D=384 \\ E=4 \\ K=9 \end{bmatrix}$* 2	$\begin{bmatrix} D=640 \\ E=4 \\ K=9 \end{bmatrix}$* 2	$\begin{bmatrix} D=768 \\ E=4 \\ K=9 \end{bmatrix}$* 2	$\begin{bmatrix} D=1024 \\ E=4 \\ K=9 \end{bmatrix}$*2
Head	1*1	Pooling and MLP	Pooling and MLP	Pooling and MLP	Pooling and MLP
Limitations (M)	10.8		27.5	51.6	93
FLOPs (B)	1.7		4.8	8.8	16.7

The proposed ViG architecture incorporates robust feature extraction techniques that are resilient to variations in neonatal appearances due to physical changes. This may include the use of deep learning models that can learn and extract discriminative features from neonatal images, even in the presence of subtle changes. The ViG architecture is trained on a diverse and representative dataset that encompasses a wide range of neonatal appearances, taking into account variations in skin color, facial features, and other relevant factors.

4. Results

Diverse dataset for neonatal switching identification was highly sought after. To gather datasets for this purpose, there have been some attempts, though the collection and use of datasets for neonatal swapping identification must take into account a number of ethical and legal considerations, including informed permission and privacy protection. Before being used in a real-world environment, models developed using these datasets should be rigorously examined for potential biases and correctness. This dataset contains pictures of babies taken in the NICU at Al-Elwiya Maternity Teaching Hospital in Al Rusafa, Baghdad, Iraq [28]. Because this is an obstetrics and gynecological hospital, all infants are regarded as aseptic. This information includes photos of healthy infants taken at various angles and under various lighting conditions. As a result, gathering as many photographs as you can further improves the accuracy. We must regularly monitor and evaluate the performance and impact of the ViG architecture to identify any ethical concerns that may arise over time. Considerations provide a

general guideline, and the specific ethical requirements may vary depending on local laws, regulations, and institutional policies. Engaging with ethicists, legal experts, and stakeholders throughout the research process can further enhance the ethical framework surrounding the ViG architecture.

Figure 6 denotes some samples that were gathered from roughly 600 newborns with a resolution of 1000X1000, all in jpg format. An iPhone 11 Pro max 12 MP camera was used to capture the pictures. The format includes the status and values for the RGB and YCrCb channels.

Figure 7 defines the loss while in training and testing, when the number iteration first increases, followed by a loss or decrease. Initially, the loss was profound and continued to reduce when the number of iterations increased. For example, at 50 iterations, loss would be 3.3%, which went to 300 iterations, where loss would be less than 1.5%.

Figure 8 describes that the computation time is essential. If the computation time is small, then this process is more advantageous. While comparing with MLP, GNN, CNN with vision GNN shows good computation time outcomes [29]. Table 3 defines the comparison of processing time with proposed model in detail.

Figure 6. Samples of Neonatal from Dataset.

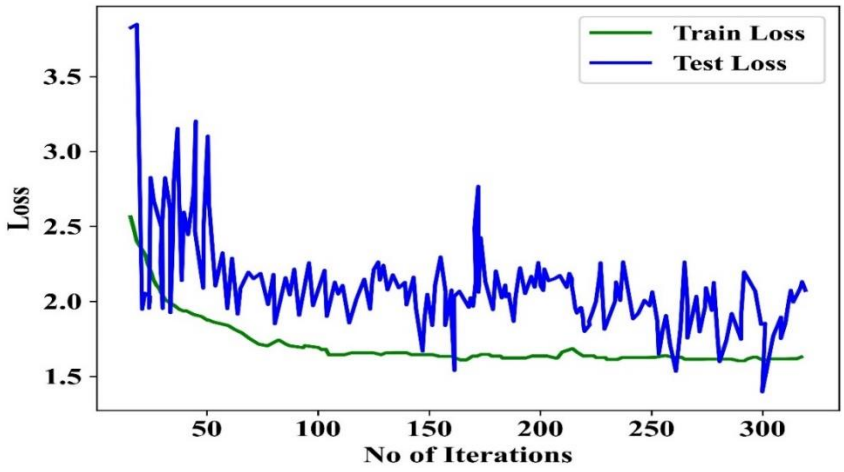

Figure 7. Training and Testing Loss.

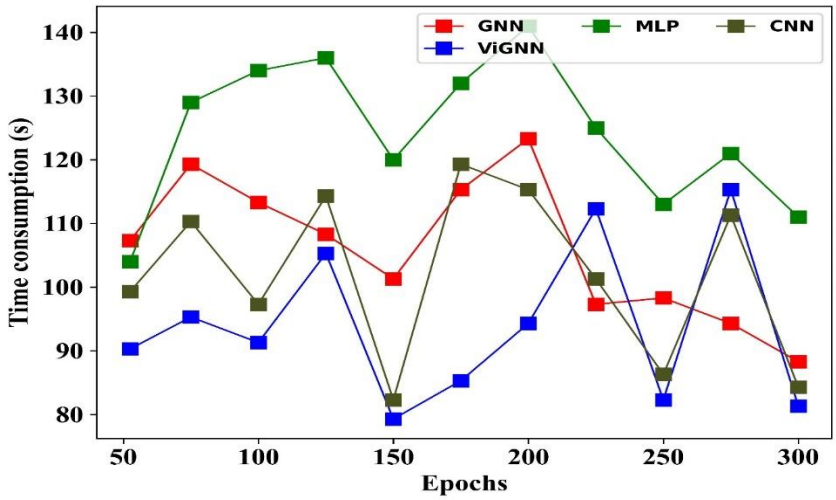

Figure 8. Computation Time for VGNN, GNN, CNN, MLP.

Table 3. Time consumption in different epochs.

EPOCH	V-GNN (In Sec)	CNN (In Sec)	GNN (In Sec)	MLP (In Sec)
50	90.3	99.3	107.3	104
100	91.3	97.3	113.3	134
150	79.3	82.3	101.3	120
200	94.3	115.3	123.3	141
250	82.3	86.3	98.3	113
300	81.3	84.3	88.3	111

Figure 9 shows the accuracy of this model in training and testing, producing a perfect level. Compared to this data set, vision GNN performs well in accuracy. During the initial stage of training the accuracy is small when the epochs increase, and the proposed model learned enough to produce an improved accuracy.

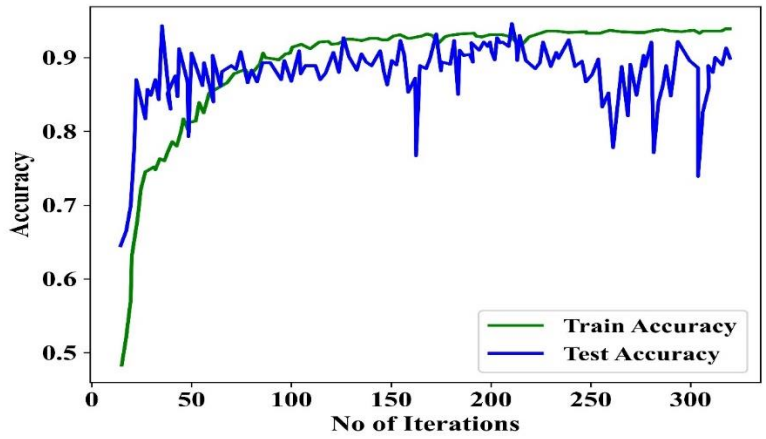

Figure 9. Accuracy analysis for the number of iterations for vision GNN.

Figure 10 illustrates the MSE, or mean squared error, which is a metric for statistical model error. It analyses the square root of the average difference between the expected and observed values. If the model had no mistakes, the MSE was equal to 0. As the model error increases, so does its value. Additionally, the MSE is sometimes referred to as the mean square deviation (MSD).

Figure 11 illustrates a confusion matrix for 150 epochs of babies taken from 10 different mothers (classes), with an accuracy of .9218.

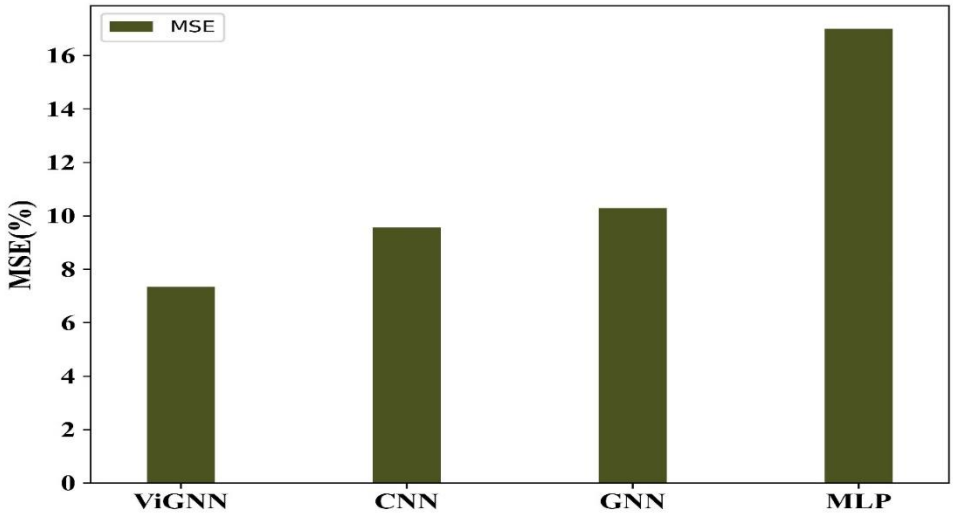

Figure 10. MSE Comparison.

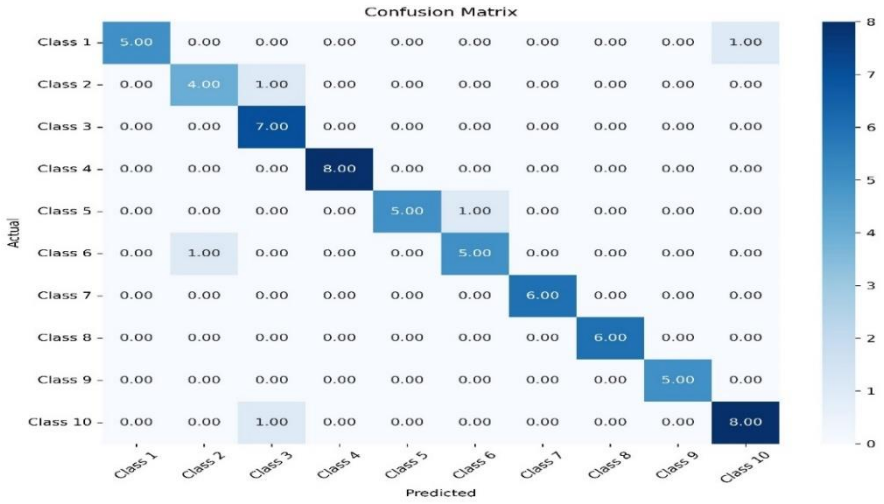

Figure 11. Confusion Matrix for 150 Epochs.

Figure 12 illustrates a confusion matrix for 300 epochs of babies taken from 10 different mothers (classes), with an accuracy of .9375.

Figure 13 illustrates the receiver operating characteristic curve/ area under the ROC Curve received a ViGNN of 95.3%, a CNN of 93.1%, a GNN of 90.0%, and an MLP of 88%.

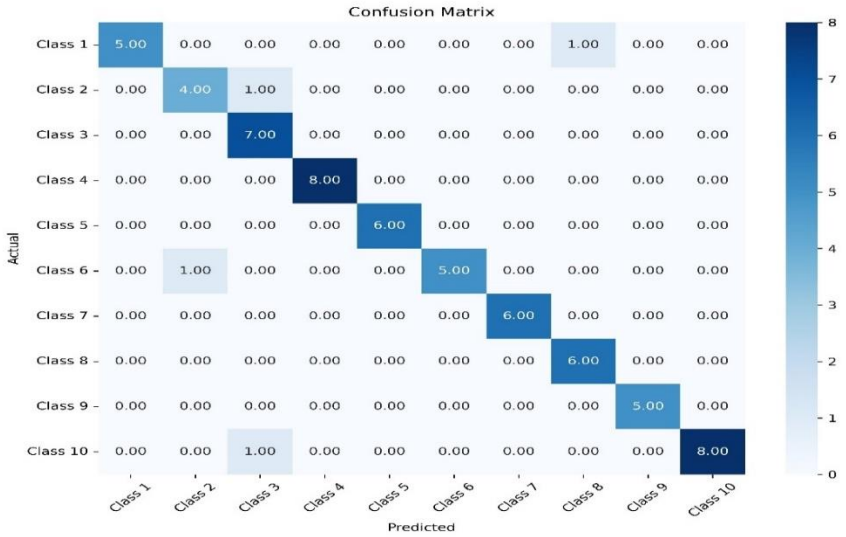

Figure 12. Confusion Matrix for 300 Epochs.

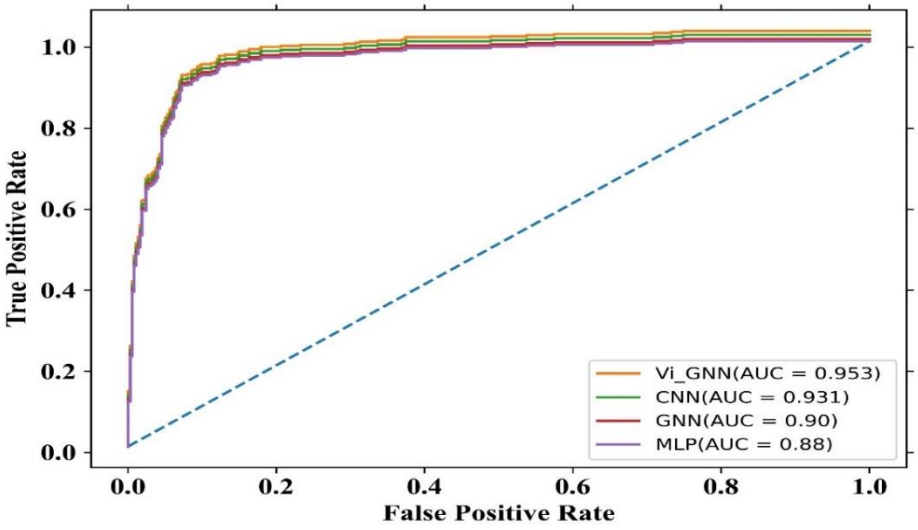

Figure 13. ROC/AUC Curve.

5. Discussions

Table 4 lists the hyperparameters utilized for the proposed vision GNN.

As seen in Figure 14, the Vision GNN had an accuracy of 92.65%, precision of 92.80%, F1 score of 92.27%, recall value of 92.25% and specificity of 98.59%. In a confusion matrix, the rate is a measurement factor, which includes four types: TPR (True positive rate), FPR (False positive rate), TNR (True negative rate), and FNR (False negative rate). The Error rate for vision GNN is 7.35, as displayed in Table 5.

Table 4. Hyperparameters.

Hyperparameters	
Batch Size	128
Learning Rate	0.0001
optimizer	Aadam
Epoch	300
Loss	Spare categorical cross Entropy
Classification layer	Softmax
No of hidden Layers	10

Table 5. Comparison of Accuracy, Recall, Precision, F1-Score, Specificity.

Algorithms	Accuracy	Precision	F1-score	Recall	Specificity	FPR	MSE
Vision GNN	93.75	94.80	93.27	93.25	98.59	0.0061	6.349
CNN	90.44	90.49	90.14	90.56	98.40	0.008	9.56
GNN	89.71	90.05	89.22	89.60	98.34	0.0086	10.29
MLP	83	81.90	81.30	82.90	88.8	0.01	17.01

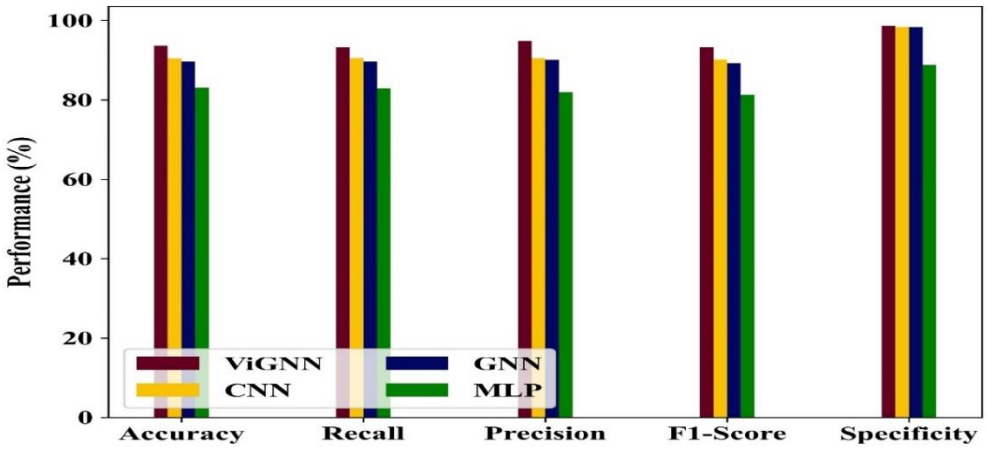

Figure 14. Comparison of Accuracy, Recall, Precision, F1-Score, Specificity.

A GNN is trained using the built graph. Deep learning models known as vision GNNs were expressly created to handle data that have a graphical structure. In order to complete tasks or generate predictions, the vision GNN learned to encode and process the data present in the graph. Neonatal identification can be performed using the vision GNN after it was trained. Their visual information is recorded and pre-processed when a new neonate arrives or needs to be identified. The input is subsequently fed into the trained vision GNN, which performs the pre-processing based on the ingested graphical representation. The vision GNN evaluates the new neonate's characteristics against the current graphical representation. The system issues an alert for potential swapping or abduction if there are any substantial anomalies or inconsistencies, such as a mismatch between the anticipated identification and the expected identity based on prior data. The technology has the ability to warn

hospital workers or security officers in the event of a suspected swapping or detected abduction. In order to avoid any harm or confusion, this facilitates a prompt response.

6. Conclusions

In this proposed work, in order to analyze the visual characteristics of newborns, such as their facial features, skin tone, and other physical characteristics, vision GNNs can be used in neonatal swapping and abduction identification. In this method, each infant is represented as a graph, where the nodes reflect various features and the edges link related features. Then, the vision GNN can learn to propagate information throughout the graph and extract features from each node, enabling it to gather the overall information of the newborn's features. The model can be trained on a large dataset of neonatal images used to identify the newborn. In the future, we will build a cognitive agent in the neonatal ICU to monitor and learn from the environment and give notifications to the authorities to avoid swapping and abduction.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

The authors extend their appreciation to the Deanship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through project number: IFP22UQU4281768DSR145.

Author contributions

Conceptualization, Y.A., and M.N.; methodology, S.R.; software, M.N.; validation, Y A., S.R. and M.N.; formal analysis, Y.A.; investigation, S.R.; resources, Y.A.; data curation, M.N.; writing—original draft preparation, S.R.; writing—review and editing, Y.A.; visualization, Y.A.; supervision, M.N.; project administration, S.R.; funding acquisition, Y.A. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflict of interest.

Reference

1. L. Hug, M. Alexander, D. You, L. Alkema, National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: Asystematic analysis, *Lancet Glob. Health*, **7** (2019), 710–720. https://doi.org/10.1016/S2214-109X(19)30163-9

2. Ø. Meinich-Bache, S. L. Austnes, K. Engan, Activity recognition from newborn resuscitation videos, *IEEE J. Biomed. Health,* **24** (2020). https://doi.org/10.1109/JBHI.2020.2978252
3. Ø. Meinich-Bache, K. Engan, I. Austvoll, Object detection during newborn resuscitation activities, *IEEE J. Biomed. Health*, **24** (2020). https://doi.org/10.1109/JBHI.2019.2924808
4. C. Skåre, A. M. Boldingh, J. Kramer-Johansen, T. E. Calisch, Nakstad, B. Nadkarni, et al., Video performance-debriefings and ventilation-refreshers improve quality of neonatal resuscitation, *Resuscitation*, **132** (2018)**,** 140–146. https://doi.org/10.1016/j.resuscitation.2018.07.013
5. S. Deepthi, P. S. Arun, Recognition of newborn babies using Multi class SVM, 2017 international conference on circuit's power and computing technologies, 2017. https://doi.org/10.1109/ICCPCT.2017.8074303
6. T. Tamilvizhi, B. Parvatha Varthini, Online vaccines and immunizations service based on resource management techniques in cloud computing, *Biomedical Research-India*; Special Issue, Special Section: Health Science and Bio Convergence Technology: Edition-I, S392-S399, 2016.
7. W. P. Jaronde, N. A. Muratkar, P. P. Bhoyar, S. J. Gaikwad, R. B. Nagrale, Review on biometric security system for newborn baby, *Int. J. Sci. Res. Sci. Technol.*, **4** (2018), 907–909.
8. S. Saurav, R. Saini, S. Singh, Facial expression recognition using dynamic local ternary patterns with kernel extreme learning machine classifier, *IEEE Access*, **9** (2021), 120844–120868. https://doi.org/10.1109/ACCESS.2021.3108029
9. S. S. Mahdi, N. Nauwelaers, P. Joris, G. Bouritsas, S. Gong, S. Walsh, et al., Matching 3D facial shape to demographic properties by geometric metric learning: A part-based approach, *IEEE T. Biometrics, Behavior, Identity Sci.*, **4** (2022), 163–172. https://doi.org/10.1109/TBIOM.2021.3092564
10. J. Zhang, G. Sun, K. Zheng, S. Mazhar, X. Fu, Y. Li, SSGNN: A Macro and microfacial expression recognition graph neural network combining spatial and spectral domain features, *IEEE T. Hum-Mach. Syst.*, **52** (2022), 747–760. https://doi.org/10.1109/THMS.2022.3163211
11. K. Han, Y. Wang, J. Guo, Y. Tang, E. Wu, Vision GNN: An image is worth graph of nodes, 36th Conference on Neural Information Processing Systems (NeurIPS 2022), 2022.
12. Z. Fu, J. Jiao, M. Suttie, J. Alison Noble, Facial anatomical landmark detection using regularized transfer learning with application to fetal alcohol syndrome recognition, *IEEE J. Biomed. Health*, **26** (2022). https://doi.org/10.1109/JBHI.2021.3110680
13. Y. Zhang, I. W. Tsang, J. Li, P. Liu, X. Lu, X. Yu, Face hallucination with finishing touches, *IEEE T. Image Process.*, **30** (2021), 1728–1743. https://doi.org/10.1109/TIP.2020.3046918
14. B. T. Susam, N. T. Riek, M. Akcakaya, X. Xu, V. R. de Sa, H. Nezamfar, et al., Automated pain assessment in children using electrodermal activity and video data fusion via machine learning, *IEEE T. Bio-Med. Eng.,* **69** (2022). https://doi.org/10.1109/TBME.2021.3096137
15. K. Michael, R. Abbas, P. Jayashree, R. J. Bandara, A. Aloudat, Biometrics and AI bias, *IEEE T. Technol. Society*, **3** (2022), 2–8. https://doi.org/10.1109/TTS.2022.3156405
16. Q. Lin, Z. Man, Y. Cao, H. Wang, Automated classification of whole-body SPECT bone scan images with VGG-based deep networks, *Int. Arab J. Inf. Techn.*, **20** (2023), 1–8. https://doi.org/10.34028/iajit/20/1/1
17. B. Ameer, A. Abdul-Hassan, VoxCeleb1: Speaker age-group classification using probabilistic neural network, *Int. Arab J. Inf. Techn.*, **19** (2022). https://doi.org/10.34028/iajit/19/6/2
18. K. A. Ogudo, R. Surendran, O. I. Khalaf, Optimal artificial intelligence-based automated skin lesion detection and classification model, *Comput. Syst. Sci. Eng.,* 44 (2023), 693–707. https://doi.org/10.32604/csse.2023.024154

19. N. Madhusundar, R. Surendran, Neonatal jaundice identification over the face and sclera using graph neural networks, Proceedings-5th International Conference on Smart Systems and Inventive Technology, ICSSIT 2023, 2023, 1243–1249. https://doi.org/10.1109/ICSSIT55814.2023.10060877
20. N. Krishnaraj, S. Rajendran, Y. Alotaibi, Trust aware multi-objective metaheuristic optimization based secure route planning technique for cluster-based IoT environment, *IEEE Access*, **10** (2022), 112686–112694. https://doi.org/10.1109/ACCESS.2022.3211971
21. S. Rajagopal, T. Thanarajan, Y. Alotaibi, S. Alghamdi, Brain tumour: Hybrid feature extraction based on UNET and 3DCNN, *Comput. Syst. Sci. Eng.*, **45** (2023), 2093–2109. https://doi.org/10.32604/csse.2023.032488
22. Y. A. Alotaibi, New meta-heuristics data clustering algorithm based on tabu search and adaptive search memory, *Symmetry,* **14** (2022), 623. https://doi.org/10.3390/sym14030623
23. S. S. Rawat, S. Singh, Y. Alotaibi, S. Alghamdi, G. Kumar, Infrared target-background separation based on weighted nuclear norm minimization and robust principal component analysis, *Mathematics,* **10** (2022), 2829. https://doi.org/10.3390/math10162829
24. R. Meenakshi, R. Ponnusamy, S. Alghamdi, O. Ibrahim Khalaf, Y. Alotaibi, Development of a mobile app to support the mobility of visually impaired people, *Comput. Mater. Con*., **73** (2022), 3473–3495. https://doi.org/10.32604/cmc.2022.028540
25. T. Tamilvizhi, R. Surendran, K. Anbazhagan, K. Rajkumar, Quantum behaved particle swarm optimization-based deep transfer learning model for sugarcane leaf disease detection and classification, *Math. Probl. Eng.,* **2022** (2022), 3452413. https://doi.org/10.1155/2022/3452413
26. Z. Dong, X. Ji, G. Zhou, M. Gao, D. Qi, Multimodal neuromorphic sensory-processing system with memristor circuits for smart home applications, 22539749. https://doi.org/10.1109/TIA.2022.3188749
27. X. Ji, Z. Dong, Y. Han, C. Lai, G. Zhou, EMSN: An energy-efficient memristive sequencer network for human emotion classification in mental health monitoring. https://doi.org/10.1109/TCE.2023.3263672
28. https://drive.google.com/file/d/16_o5NU1GDmAS85lkAg-9fBxvy2Du67Vf/view
29. T. Thanarajan, Y. Alotaibi, S. Rajendran, K. Nagappan, Improved wolf swarm optimization with deep-learning-based movement analysis and self-regulated human activity recognition, *AIMS Math.*, **8** (2023), 12520–12539. https://doi.org/10.3934/math.2023629

AIMS Press