DOI: https://doi.org/10.52756/ijerr.2024.v38.017 *Int. J. Exp. Res. Rev.*, **Vol. 38: 182-193 (2024)**

Original Article | *Peer Reviewed* | *Open Access*

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in
ISSN 2455-4855

Deep Learning-Based Architecture for Down Syndrome Assessment During Early Pregnancy Using Fetal Ultrasound Images

Aijaz Ahmad Reshi[1]*, Shabana Shafi[1], Irfan Qayoom[2], Maria Wani[3], Shahida Parveen[4] and Ajaz Ahmad[5]

[1]Department of Computer Science, College of Computer Science and Engineering, Taibah University, Al Madinah Al Munawarah, Saudi Arabia; [2]Neonatal Intensive Care Unit, Department of Pediatrics, King Faisal Specialist Hospital and Research Centre, Madinah, Saudi Arabia; [3]Department of Radiodiagnosis, Sher-I-Kashmir Institute of Medical Sciences, Srinagar, Jammu and Kashmir, India; [4]Department of Nursing, College of Pharmacy and Applied Medical Sciences, Dar Al Uloom University, Riyadh, Saudi Arabia; [5]Department of Clinical Pharmacy, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia

E-mail/Orcid Id:

AAR, aijazonnet@gmail.com, https://orcid.org/0000-0002-0795-2835; *SS,* bhatshabu@gmail.com; *IQ,* premieneonate24@gmail.com; *MW,* altafwani1958@gmail.com; *SP,* sparveen@dau.edu.sa; *AA,* ajukash@gmail.com, https://orcid.org/0000-0001-6295-8685

Article History:

Received: 12th Feb., 2024
Accepted: 17th Apr., 2024
Published: 30th Apr., 2024

Keywords:
Trisomy 21, Down syndrome, Artificial Intelligence, Convolutional Neural Network, Nuchal Translucency

How to cite this Article:
Aijaz Ahmad Reshi, Shabana Shafi, Irfan Qayoom, Maria Wani, Shahida Parveen and Ajaz Ahmad (2024). Deep Learning based architecture for Down syndrome assessment during early pregnancy using fetal ultrasound images. *International Journal of Experimental Research and Review*, *38*, 182-193.
DOI:
https://doi.org/10.52756/ijerr.2024.v38.017

Abstract: Down's syndrome, also known as trisomy 21, is a prevalent genetic condition characterized by intellectual disability and developmental delay in children. The current study has prioritized the development of precise screening techniques for trisomy 21 in the initial stages of pregnancy in order to facilitate prompt diagnosis. This study presents an innovative paradigm for categorizing sagittal views in obstetric ultrasound examinations, specifically identifying Nuchal Translucency, a crucial component within the fetal brain, during the 11th to 14th weeks of pregnancy. The suggested deep learning-based system effectively detects the presence of the essential cerebral structure known as Nuchal Translucency, hence aiding in the diagnosis of Down's syndrome. A dataset comprising more than 1100 pre-processed 2D sagittal-view ultrasound images was gathered to train, test, and validate the proposed convolutional neural network model. The model results were utilized to quantify neurotensin levels and assess the presence of Down's syndrome by image classification. The performance of the model was assessed by measuring its sensitivity, specificity, and area under the curve metrics. These metrics were then compared to those of human experts who had received training in prenatal and ultrasound techniques. Notably, the suggested model attained an outstanding area under the curve score of 0.97. Our study suggests the most common non-invasive method for screening pregnant women for fetal abnormalities is ultrasound, which examines the unborn organs. The application of deep learning and machine learning techniques has significantly enhanced the diagnosis. In order to detect or anticipate conditions like Down syndrome, it assesses the intercranial structures of the developing embryo during the early stages of pregnancy. Developing an improved iteration of this model could serve as the foundation for an effective automated system for diagnosing Down's syndrome in its early stages within a clinical environment.

Introduction

Individuals with Down's syndrome are believed to have always existed throughout civilization. However, the medical recognition of this disorder did not occur until the 19th century. However, it was not until the year 1866 that the English physician John Langdon Down released a detailed account of the condition that would subsequently be named after him (Ataman et al., 2012). Trisomy 21, characterized by the presence of an additional chromosome 21, results in various clinical abnormalities known as Down syndrome (Akhtar and Bokhari, 2023). Down syndrome (DS) is the most

***Corresponding Author:** aijazonnet@gmail.com

frequently occurring autosomal aneuploidy that may be survived, and it is a genetically intricate condition that is linked to the ability of humans to survive beyond the normal term of pregnancy (Antonarakis et al., 2020; Shafi et al., 2024). The disorder frequently results in many anomalies in individuals, including intellectual incapacity and hereditary predisposition to congenital heart illness (Antonarakis et al., 2020; Ataman et al., 2012). Down syndrome is the most common autosomal aneuploidy, and it is also the most common genetic disorder in neonates. The sentinel trait is easily identifiable and may be verified using cytogenetic analysis (Akhtar and Bokhari, 2023). Trisomy 21 is a well-known example of aneuploid mutations, which are extensively recorded in various congenital malformation monitoring systems. One such system is the European Surveillance of Congenital Anomalies (EUROCAT), a network of approximately 40 registries based on population data, established in 1979 (Sperling et al., 2023). During the last thirty years, there have been notable progressions in the diagnosis and prenatal identification of Down syndrome. From 1990 to 2021, antenatal screening services have become more readily available, allowing pregnant women to acquire more accurate information about their pregnancies at an early stage without having to undergo invasive procedures such as Chorionic villus sampling (CVS) or amniocentesis (Alfirevic and Neilson, 2004; Russo and Blakemore, 2014). Consequently, antenatal screening became mandatory for the detection of Down syndrome in numerous countries. As of 2021, a minimum of 76.9% of countries offered complete government funding for DS diagnostic tests (Sperling et al., 2023).

The risk of Down syndrome in the developing fetus is assessed through ultrasound imaging and a blood test usually between the eleventh and thirteenth weeks of pregnancy (Crossley et al., 2002; Li et al., 2022). The nuchal translucency (NT) is a fluid-filled area at the back of the fetal neck that is measured as part of this evaluation together with the findings of the blood test and the mother's age (Dominic-Gabriel and Roxana-Cristina, 2018). By taking a holistic view, we may detect an increased risk of numerous additional, less frequent, chromosomal abnormalities and, with an accuracy of 85 to 90%, predict the probability of Down syndrome. Ultrasounds can also help find abnormalities or structural problems with the growing baby (Dominic-Gabriel and Roxana-Cristina, 2018). There has been a lot of interest in using AI and ML techniques for medical image analysis to diagnose diseases, thanks to the encouraging findings in the past decade (Barragan-Montero et al., 2021; Kumar et al., 2023; Bulawit et al., 2023). In recent years, convolutional neural networks (CNNs) and other Deep Learning (DL) models have shown remarkable performance when it comes to quantitatively analyzing different kinds of medical images. Images from X-ray, CT, and MRI scans are some examples of what can be classified (Sarvamangala and Kulkarni, 2022; Yousef et al., 2022; Haloi et al., 2023). Using CNNs for US image processing has also demonstrated remarkable outcomes in numerous previous research projects, particularly in identifying and diagnosing a variety of ailments (Sarvamangala and Kulkarni, 2022; Kaur, 2023; Reddy and Khanaa, 2023). The convolutional and subsampling sub-layers that make up a complete convolutional neural network (CNN) architecture are shown in Figure 1. The mass of units that have the same color is equal. Any amount of fully connected layers can be added to the network topology after the convolutional layers. The question that remains for the DL method is whether or not it can successfully automate the capture of the desired features, making it an accurate tool for fetal DS screening. The primary goal of this research is to find out if DS may be diagnosed in early pregnancy using a DL-based noninvasive approach that can identify fetal intracranial characteristics. As a result, the following goals have motivated the development of a DL-based model for the accurate classification of sagittal views in obstetric ultrasound images: In order to find out if AI can diagnose fetal intracranial abnormalities in the first trimester of pregnancy. To provide a method for automatically categorizing sagittal views in obstetric ultrasound images. To suggest a new CNN design for US image classification using NT. In order to help health practitioners, weigh the benefits and risks of sophisticated invasive exams before deciding whether to proceed with treatment.

Materials and Methods

Dataset

This study's dataset includes 1,120 2D ultrasound images collected from various medical facilities (Chen and Fajin, 2022). The images in the dataset are of fetuses that are 11–14 weeks along in the pregnancy. The initial image dataset was optimized to train the suggested CNN by applying image processing and computer vision techniques. The methods used for data preparation, training the models, and assessing their performance are detailed in the sections that follow.

Feature Extraction Phase of the model

Figure 1 shows that the feature extraction stage is critical in CNN. In this step, we look for important aspects of an image and pull them out. This phase is

DOI: https://doi.org/10.52756/ijerr.2024.v38.017

crucial since the characteristics will be utilized to detect image anomalies. In CNNs, a convolution layer is an essential building block for feature extraction. It is common for it to incorporate both linear and nonlinear processes, such as activation functions and convolution (Yamashita et al., 2018). Mathematically, the operations in a convolutional layer can be expressed as follows:

Let X represent the input image with dimensions m×m×r.

K be the number of filters (kernels) in the convolutional layer.

n be the size of each filter (n×n×q), where

q is either equal to or less than r.

P denotes the pooling size (p×p).

$$F_i = \sigma\left(\sum_{j=1}^{q}\sum_{u=1}^{n}\sum_{v=1}^{n} W_{i,j,u,v} * X_{j,u+x-1,v+y-1}\right) + b_i \quad (1)$$

Where ***Fi*** is the ith feature map, ***σ*** is the activation function, (Rectified Linear Unit activation function (RELU in this case)) $W_{i,j,u,v}$ are the weights $X_{j,u,v}$ is the input image patch and bi is the bias.

preventing overfitting during training. Pooling operation can be expressed mathematically as:

$$F_i' = F_i \downarrow P \quad (2)$$

Where F'_i is the subsamped feature map and ↓**P** represents mean or maximum pooling over ***p x p*** contiguous portions.

Fully connected Layer

We can extract visual features using the convolution layer, which gets pooled and down-sampled before becoming the network's final output. Be advised that in a CNN, the number of clusters is directly proportional to the number of nodes contained in each fully connected layer. This ensures that a human observer can understand all the information relevant to a specific job, like clustering because it is all contained within a single layer. The operations of the layer can be represented as:

$$Y_k = f\left(\sum_{i=1}^{K} W_{k,i} \cdot F_i'\right) \quad (3)$$

Where Y_k is the output of the kth neuron in the fully connected layer, Wk and i are the weights, and ***f(.)*** is the activation function.

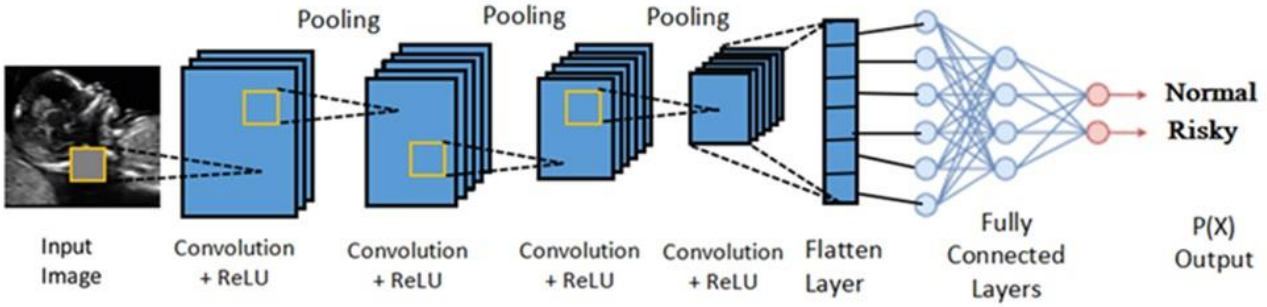

Figure 1. CNN general architecture.

Pooling or Subsampling Operation

Pooling or subsampling is a fundamental operation in convolutional neural networks (CNNs) that is used to downsample the spatial dimensions of feature maps. This operation helps reduce computational complexity and extract the most relevant information. Typically, max pooling or average pooling is employed. In max pooling, the feature map is divided into non-overlapping regions, and the maximum value within each region is retained, discarding the rest. This retains the most salient features. A similar process is followed in average pooling, but the average value within each region is calculated. It provides a smoothed-down version of the features. Both max pooling and average pooling contribute to making the CNN translationally invariant by focusing on the most essential features while reducing the dimensionality of the data, aiding in efficient computation and

Probability Calculation

A wide variety of activation functions can be used to create the last completely connected layer of a neural network. While other kinds of functions may be better suited to certain issues, the linear function is by far the most common. To normalize output values and offer target class probabilities, CNNs commonly use the softmax function. Here, each value ranges from 0 to 1, and all of them add up to 1. For the last layer, most people choose an activation function that is either linear or sigmoid.

Performance parameters

The goal is to determine whether a CNN is good enough to implement by evaluating and validating its performance. AUC of the Receiver Operating Characteristic (ROC) is computed as part of the evaluation technique. The area under the receiver

DOI: https://doi.org/10.52756/ijerr.2024.v38.017

operating characteristic (AUC-ROC) curve is useful for visualizing CNN classifier performance. Even though it's limited to binary classification problems, it can be modified to assess problems with more than two classes. Sensitivity and specificity are the two main performance metrics that can be utilized to construct the AUC-ROC curve. The performance measures utilized to evaluate this study's proposed CNN model architecture are explained in the following subsections.

Sensitivity

A CNN model's sensitivity indicates how effectively it can detect positive examples. The terms recall and true positive rate (TPR) are synonyms for sensitivity. One way to measure a model's efficacy is by looking at its sensitivity, which reveals the proportion of positive instances that it correctly identified. A CNN model with high sensitivity means that it is missing classifying very few positive instances to classify them as positive (Reshi et al., 2021). So, high sensitivity implies less false negatives. The use cases covered in this study make it crucial for the model to have high sensitivity as we are bound to all true positive instances. The measure can be calculated as:

Sensitivity = (True Positive) / (True Positive + False Negative) (1)

True Positive

Images that are anticipated to be risky (or with the possibility of Down syndrome) are actually Risky (or with the possibility of Down syndrome); in other words, the true positive is the total number of images that are both predicted to be and actually in risky class.

False Negative

Images that actually belong to the Risky class (or with the possibility of Down syndrome) are classified in the normal class. In other words, the number of images classified as normal but risky is represented by the false negative. As it may turn out to be a significant failure of the model. The model should ideally have a low false negative rate. A higher true positive score and a smaller false negative score result in higher sensitivity. A lower sensitivity would result in a higher false negative and a lower true positive value. Models with a high sensitivity will be sought for healthcare and diagnostic applications.

Specificity

It is common practice to evaluate sensitivity and specificity when evaluating the efficacy of a model. Specificity refers to the percentage of false negatives that the model accurately identifies. This adds to the number of what can be called false positives actually negative results that were mistakenly labelled as positive. The True Negative Rate (TNR) is another possible way to characterize this metric. The sum of the specificity (the rate of genuine negatives) and the false positive rate would always be 1. When it comes to negative outcomes, a specificity-rich model will catch most of them, while a less specific one can wrongly mark a lot of bad occurrences as good (Taye, 2023). Assume for the sake of argument that the model provides a 95% specificity score in order to illustrate the present study scenario. In other words, out of 100 images that should have been classed as normal, the model has accurately identified 95 as normal and labelled just 5 as dangerous, which means that there were 5 false positives.

Specificity = (True Negative)/ (True Negative + False Positive) (2)

True Negative

How many images are considered normal and probably normal? To rephrase, the true negative is the proportion of expectedly normal images that turn out to be normal.

False positives

The number of typical images labeled as potentially dangerous. Put another way, the false positive represents the number of images that were wrongly classified as dangerous when, in fact, they were normal. The model would work best with a true negative rate or extremely high specificity. A higher specificity score results in a lower false positive rate and a greater true negative score. However, more false positives and fewer real negatives are the results of a low specificity score (Rustam et al., 2020).

AUC-ROC Curve

The Receiver Operating Characteristic (ROC) curve is a visual representation of the performance of a binary classification model, commonly using metrics such as sensitivity (true positive rate) and specificity (true negative rate). The evaluation of model performance extends to the calculation of the Area Under the ROC Curve (AUC), a numerical measure providing an overall assessment.

I) *ROC Curve:* The ROC curve is a parameterized plot depicting TPR against FPR for diverse classification thresholds. Mathematically, it is expressed as:

$$\mathrm{ROC}(t) = (FPR(t), TPR(t)) \quad (4)$$

Where, ***t*** signifies the classification threshold.

II) *Area Under the ROC Curve (AUC):* The AUC is determined by integrating the ROC curve. It represents the probability that the model ranks a randomly selected positive instance higher than a randomly chosen negative instance. Mathematically, AUC is given by:

DOI: https://doi.org/10.52756/ijerr.2024.v38.017

$$AUC = \int_0^1 TPR(FPR^{-1}(t))\,dt \quad (5)$$

where FPR−1 denotes the inverse function of FPR.

The AUC ranges between 0 and 1, with a higher value indicating superior model performance. A value of 0.5 implies a model performing no better than random chance. In summary, the ROC curve visually illustrates the trade-off between sensitivity and specificity at different classification thresholds, while the AUC provides a quantitative measure of the model's overall discriminatory power. Since the sensitivity and specificity are used to plot the ROC curve. The area under the receiver operating characteristic curve (AUC) is also used to evaluate the model's performance. Figure 2 shows three examples of ROC curves along with the corresponding AOC values.

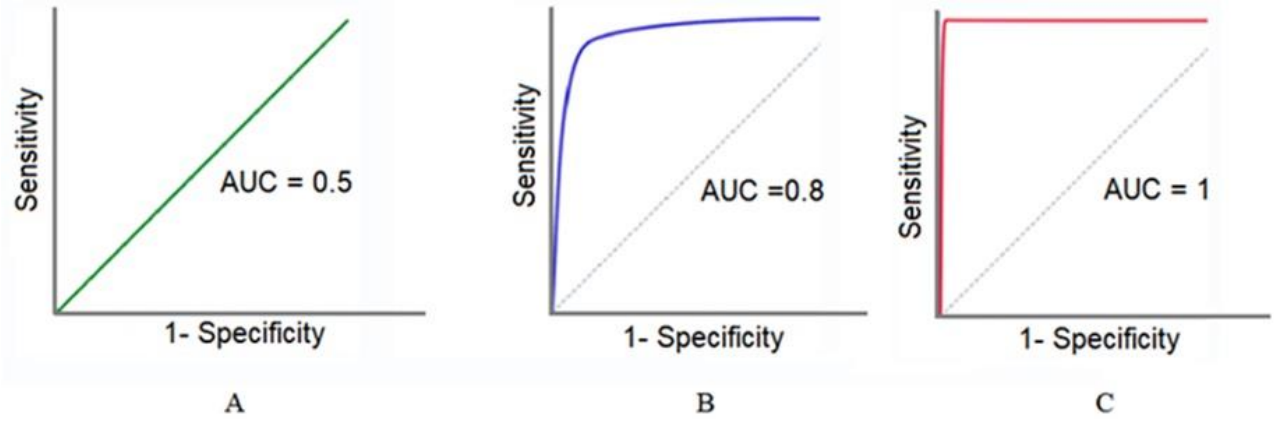

Figure 2. ROC-AUC Curve A) AUC = 0.5 B) AUC = 0.8 and C) AUC = 1.

Results

Image segmentation

One of the most crucial steps in medical image preparation for analysis across modalities is image segmentation to extract the ROI. The section of a 2D image that is most suitable for assigning a specific class is called the ROI, or subset of the image. Expert humans are typically enlisted to aid in the ROI extraction process. Lately, ML and DL methods have considerably simplified this hitherto laborious procedure; these methods can handle and analyze massive amounts of images to extract structural information. The processing and analysis of entire images is impossible with modern computers' CPU and memory capacity. However, by analyzing only the intended ROIs, the performance and speed of the analysis algorithms are much improved. We used ROI extraction to remove the parts of the US image that aren't relevant to NT classification but could have hampered the suggested model's efficiency and effectiveness. For the purpose of training the model, each image from the original dataset had its ROI containing the NT region excised. A bounding box has been used to segment each US image based on the features extracted from the item detection method. The ROI's pixel coordinates of the upper left and lower right corners were included in the object detection features. Table 1 shows the object detection feature set sample. There are examples of both the original and segmented images in figure 3. They were all cross-checked to ensure that the desired region was captured in each ROI-based image. The described ROI extraction process involves segmenting ultrasound (US) images using bounding boxes obtained from object detection features. Let's represent this mathematically:

Let, I be the original US image.

B is the bounding box obtained from the object detection features.

ROI be the extracted Region of Interest.

The bounding box

B is defined by the pixel coordinates of its upper-left corner and lower-right corner represented as follow:

$$(x_{\text{upper-left}}, y_{\text{upper-left}})$$

$$(x_{\text{lower-right}}, y_{\text{lower-right}})$$

The ROI extraction process can be represented as:

Object Detection Features (ODF): The object detection features include the pixel coordinates of the upper-left and lower-right corners of the bounding box:

$$ODF = \{(x_{upper-left}, y_{upper-left}), (x_{lower-right}, y_{lower-right})\} \quad (6)$$

Bounding Box Definition: Bounding box B is defined using the object detection features:

$$B = \{(x, y) \mid x_{upper\text{-}left} \leq x \leq x_{lower\text{-}right} , y_{upper\text{-}left} \leq y \leq y_{lower\text{-}right} \}$$

(7)

ROI Extraction: The ROI can be defined as the portion of the original image ***I*** where the coordinates ***(x,y)*** fall within the boundaries of the bounding box ***B***. This can be expressed mathematically as follows:

$$ROI = I(x, y) \mid (x, y) \in B$$

(8)

The notation indicates that the ROI comprises all pixel values I(x,y) from the original image I corresponding to coordinates (x,y) that lie inside the bounding box B. In essence, it represents the selected area of interest delineated by the bounding box.

Proposed CNN Model Architecture

With eleven layers, the suggested CNN achieves its impressive speed and accuracy. The model uses an image with dimensions of 100 by 400 pixels. A self-defined residual block allows the model to extract features at various levels. The discriminative information the model focuses on can be seen through the visualization of feature maps. Model overfitting can be prevented in the proposed design by utilizing the dropout technique. We started with a learning rate 0.0001 and a batch size of 16. Every iteration of training a model consists of 300 epochs. The model's learnable parameters were fine-tuned using the well-known Adam optimizer to increase the model's accuracy. The model has been penalized for training-related misclassification due to mistakes in probability computation using binary cross entropy. To forecast the DS risk score for every image, the model's

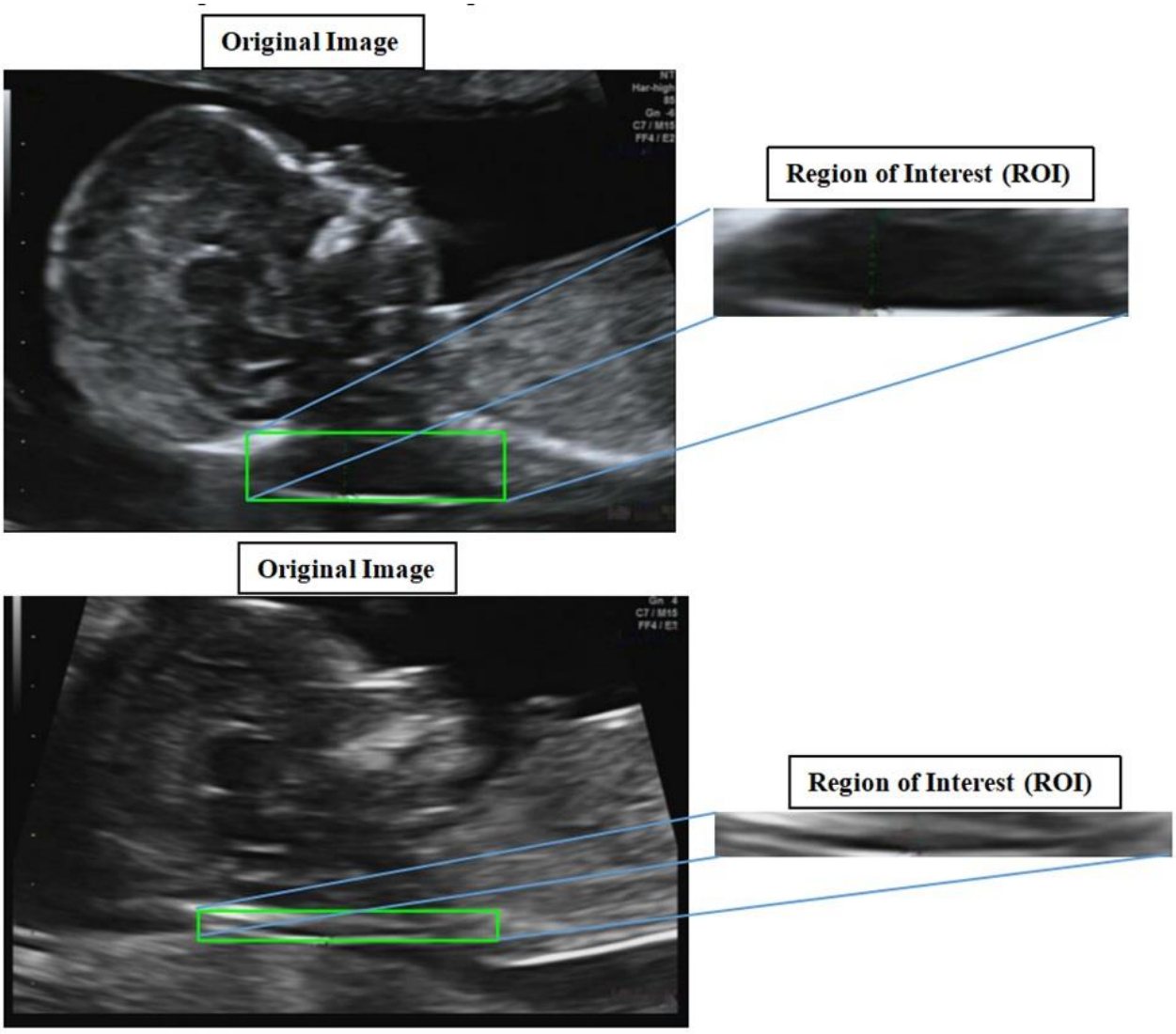

Figure 3. Original images and corresponding ROIs extraction by the proposed method.

DOI: https://doi.org/10.52756/ijerr.2024.v38.017

output likelihood that can take on values between zero and one has been defined. At last, the images are sorted into two categories. If the probability score is high, the image is considered at high risk of DS; if the score is low, the image is considered normal. Figure 4 shows the detailed architecture.

Table 1. Sample object detection feature set.

Image ID	Xmin	Ymin	Xmax	Ymax
0.png	252	376	519	446
1.png	121	390	449	422
2.png	109	380	386	409
3.png	169	393	410	414
4.png	213	357	416	393

The main functional components of the CNN are given as:

Residual Block: A residual block is an integral component that facilitates learning residual functions, streamlining the training of deep networks. Mathematically, the output of a residual block given an input x can be expressed as:

$$Output = f(x) + x \tag{9}$$

where ***f(x)*** denotes the transformation applied by the residual block.

1. Dropout Technique: Dropout serves as a regularization technique by randomly setting a fraction of input units to zero during training. Mathematically, for a given layer with input x, the output incorporating dropout can be represented as:

$$Output = dropout(X) \tag{10}$$

2. Learning Rate and Batch Size: Learning rate (α) and batch size are essential hyperparameters. The update rule for adjusting model parameters during training through gradient descent is given by:

$$\theta_{new} = \theta_{old} - \alpha \,.\, \nabla\theta L \tag{11}$$

θ_{old} : Represents the current values of the model parameters before applying any updates.

α (Learning Rate): Determines the size of the step taken during each iteration of the optimization process.

∇θL: Represents the gradient of the loss function L with respect to the model parameters θ.

θ_{new}: Denotes the updated values of the model parameters after applying the gradient descent update rule.

3. Adam Optimizer: The Adam optimizer amalgamates concepts from momentum and Root Mean Squared Propagation (RMSProp) to dynamically adjust learning rates during training. The update rule for Adam is delineated by:

$$m_t = \beta_1 . m_{t-1} + (1-\beta_1).\nabla_\theta L \tag{12}$$

$$v_t = \beta_2 . v_{t-1} + (1-\beta_2).(\nabla_\theta L)^2 \tag{13}$$

$$\theta_{new} = \theta_{old} - \frac{\alpha}{\sqrt{vt} + \varepsilon} . m_t \tag{14}$$

Where m_t and v_t denote moving averages of the gradient and its square, α is the learning rate, and ε is a small constant to prevent division by zero.

4. Binary Cross Entropy Loss: Binary Cross Entropy emerges as a widely used loss function for binary classification tasks. In a binary classification scenario with true labels y and predicted probabilities ŷ the loss is articulated as:

$$L = -\frac{1}{N} \Sigma_{i=1}^{N} y_i . \log(\hat{y}_i) + (1-y_i).\log(1-\hat{y}_i) \tag{15}$$

ROC Curve Analysis

Our suggested CNN model was trained using 1,120 images taken in the United States. With an AUC score of 0.97 for the training set and 0.94 for the validation set, the suggested CNN performs admirably on both sets of data. In Figure 5, we can see a graph that represents the presented performance metrics. The specificity (FP rate) is shown on the X-axis and the sensitivity (TP rate) on the Y-axis of the graph. The ROC curves for the training set and validation set are shown in red and green, respectively.

CNN Model Robustness

The dataset was randomly divided into three equal sections, two for training and the other for validation to test the CNN model's robustness. To examine how variable training set selection on all US images affects the proposed model. Through three-fold cross-validation, the suggested CNN's robustness was assessed. All images were arranged by name in ascending order, with two chunks of namespace representing corresponding images as the training set and the rest as the validation set. The first two halves were used as training sets, and the third as cross-validation sets. II) The first component is validation and the last two is training. III) The center part was validation, and the first and last sections were

DOI: https://doi.org/10.52756/ijerr.2024.v38.017

training. Table 2 summarizes model performance in training and validation sets and shows area under curve score graphs. Figure 6 (A), 6 (B), and 6 (C) depict the first, second, and third cross-validation scenarios. Results indicate AUCs above 0.90 for each training or validation set. The findings show that the proposed model performs effectively regardless of training and validation set splits.

Table 2. Performance of validation scenarios

Scenario I	Training set	AUC = 0.97
	Validation set	AUC =0.92
Scenario II	Training set	AUC =0.96
	Validation set	AUC =0.90
Scenario III	Training set	AUC =0.97
	Validation set	AUC =0.93

Discussion

CNNs or convents are increasingly used in computer vision and image categorization (Sarvamangala & Kulkarni, 2022). Image object identification previously required manual feature extraction. However, CNNs' scalable image classification and object recognition have revolutionized the area (Alzubaidi et al., 2021). They find image patterns using matrix multiplication and algebra. CNNs are computationally costly and often require GPUs for training (Lim et al., 2023). CNNs are easier to train and have fewer parameters than fully linked networks. CNNs start with numerous convolutional and subsampling layers before fully connected layers (Taye, 2023; Zafar et al., 2022). CNNs can classify and evaluate medical image data (Nia et al., 2023). Our study introduces a multi-layer CNN-based classification system for early Down syndrome diagnosis in prenatal care. Our model classifies 2D sagittal view ultrasonography (US) images as normal or at risk for Down syndrome using cutting-edge deep learning techniques for medical image analysis. The NT area of the US images is crucial to early Down syndrome diagnosis. We contribute to the discussion regarding using artificial intelligence, particularly deep learning, in medical image analysis for disease diagnosis. Numerous studies have shown that these technologies can alter radiology and prenatal care (Ling et al., 2019).

Many countries have Down syndrome prenatal screening programs. These programs start with a screening test and then offer a more invasive diagnostic test for 'high risk' results to confirm the diagnosis (Hill et al., 2016). Machine learning and deep learning have become popular in medical predictive analytics (Ahuja, 2019). Due to extremely uneven and feature-correlated screening data, these technologies have been limited in their ability to predict Down syndrome (Ling et al., 2019). Down syndrome diagnosis has been a major achievement in prenatal medicine in recent decades (Van den Veyver, 2016). Different methods have different detection rates, acceptability, costs, and downsides. While early genetic screening technologies have improved, first-trimester ultrasounds remain critical (Dominic-Gabriel & Roxana-Cristina, 2018). The introduction of the nuchal translucency (NT) measurement represents a significant mile-stone in first-trimester screening (Niknejadi and Haghighi, 2015; Roozbeh et al., 2017). Nuchal translucency (NT) assessment was a major advance in first-trimester screening. Between 11th and 14th weeks of gestation, a distinctive, fluid-filled region behind the fetal neck is carefully examined. A high NT value is significantly connected with fetal Down syndrome and other aneuploidies, allowing the diagnosis of 70% of Down syndrome cases with a good false-positive rate of 5% (Driscoll et al., 2008). Sonologists use calipers horizontally on NT linings to estimate NT measurement manually or semi-automatically (Nasibeh Roozbeh et al., 2017). The sonologist's talents and experience determine manual estimation. These factors can delay and miscalculate measurements (Cho et al., 2015). Considering the drawbacks of conventional NT measurement and CNNs' exceptional performance in medical image processing. We meticulously choose ROIs from US images taken throughout 10-14 weeks of pregnancy to train our CNN. This ROI, which includes the NT region, is rigorously identified using manual labeling and object detection. As mentioned, medical image databases require high data quality and specificity. As CNNs have repeatedly outperformed other deep learning methods in medical image processing, our findings support this (Snider et al., 2022). Our investigation shows that the suggested CNN architecture shines when trained on these ROIs, demonstrating its versatility and usefulness in medical image analysis for DS diagnosis. Sensitivity, specificity, and ROC-AUC curve analysis are used to objectively test our CNN model. As mentioned previously, these measures are critical for assessing the reliability and accuracy of AI models in medical diagnostics and emphasise the significance of rigorous performance evaluation (Liu et al., 2019). Besides model training, our study includes testing sets and validation situations. This fits the story of AI in medical image analysis, which acknowledges the diversity of medical image datasets and the need for rigorous validation (Boice et al., 2022). The outcomes of our study are promising and support our approach. We found that our model regularly obtains AUC values betw-

DOI: https://doi.org/10.52756/ijerr.2024.v38.017

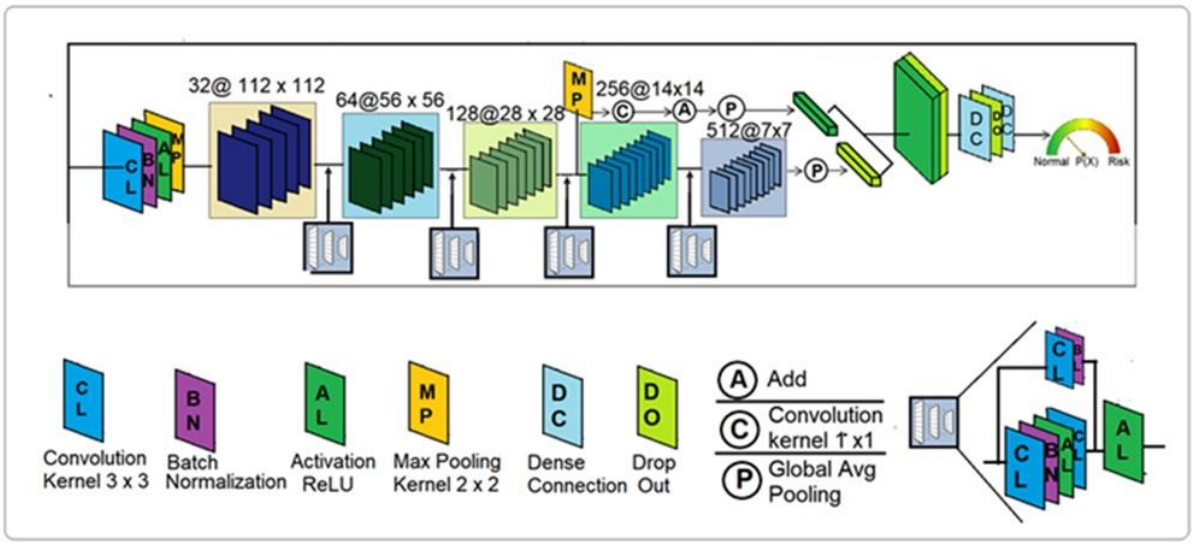

Figure 4. Detailed architectural elements and work flow of the proposed CNN.

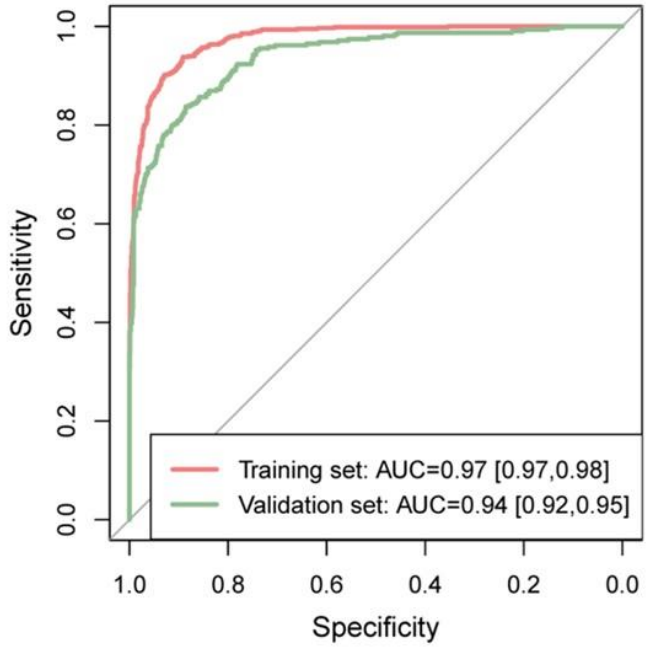

Figure 5. Area Under Curve (AUC) graph.

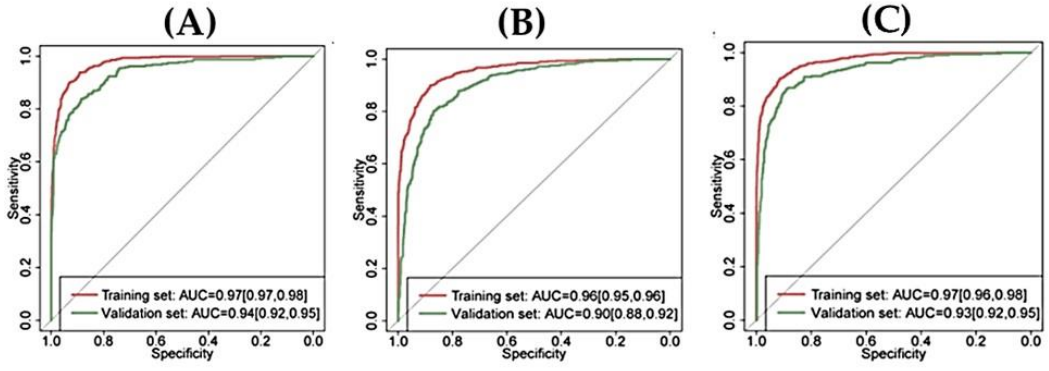

Figure 6. (A) AUC for Training and Testing for scenario I of cross-validation; (B) AUC for Training and Testing for scenario II of cross-validation (C) AUC for Training and Testing for scenario III of cross-validation.

DOI: https://doi.org/10.52756/ijerr.2024.v38.017

-een 0.90 and 0.97 in testing and validation settings. These findings support AI's involvement in medical predictive analytics and DS prediction (Liu et al., 2019; Wang et al., 2021). Our study advances prenatal by using deep learning and CNNs to better early diagnosis and care. Our model's effectiveness and reliability demonstrate its potential to address Down syndrome screening and medical image analysis concerns. We use AI to improve prenatal care and discover Down syndrome early, which has great potential to improve healthcare for expecting parents and their children.

Conclusion

Ultrasonography is the most used non-invasive pregnancy screening tool for fetal abnormalities by evaluating fetal organs. It evaluates the fetus' intercranial structures in early pregnancy to diagnose or predict diseases like Down syndrome. Predicting Down syndrome with NT testing is common and reliable. This paper introduces a CNN architecture for US image classification to estimate DS probability using NT. Predicting Down syndrome in the fetus with NT is important. The work trains the CNN model using NT-based ROI to increase performance. The planned 11-layer CNN uses 2D US images. The model extracts different feature levels using a self-defined residual block. The suggested architecture uses dropout to prevent model overfitting. The suggested CNN has good training and validation performance with AUC scores of 0.97 and 0.94, respectively. More training images are needed to improve the suggested CNN's performance. An effective automated method for early diagnosis of Down syndrome in clinical settings might start with the enhanced model implementation. Understanding varied clinical contexts and model optimizations is needed to adapt the suggested framework to diverse application scenarios.

Abbreviations

deep learning (DL); machine learning (ML); Nuchal Translucency (NT); ultrasound (US); convolutional neural network (CNN); area under the curve (AUC); Chorionic villus sampling (CVS); Receiver Operating Characteristic (ROC); True Negative Rate (TNR); region of interest (ROI); Graphics Processing Units (GPUs)

Acknowledgements

The authors extend their appreciation to the King Salman Center for Disability Research for funding this work through Research Group no KSRG-2023-562.

Funding

King Salman Center for Disability Research (KSRG-2023-562)

Conflict of Interest

None

References

Ahuja, A. S. (2019). The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ, 7*, e7702. https://doi.org/10.7717/peerj.7702

Akhtar, F., & Bokhari, S. R. A. (2023). Down Syndrome. [Updated 2023 Aug 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK526016/.

Alfirevic, Z., & Neilson, J. P. (2004). Antenatal screening for Down's syndrome. *BMJ, 329*(7470), 811-812. https://doi.org/10.1136/bmj.329.7470.811

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., . . . Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data, 8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

Antonarakis, S. E., Skotko, B. G., Rafii, M. S., Strydom, A., Pape, S. E., Bianchi, D. W., . . . Reeves, R. H. (2020). Down syndrome. *Nat Rev Dis Primers, 6*(1), 9. https://doi.org/10.1038/s41572-019-0143-7

Ataman, A. D., Vatanoglu-Lutz, E. E., & Yildirim, G. (2012). Medicine in stamps: history of Down syndrome through philately. *J. Turk. Ger. Gynecol. Assoc., 13*(4), 267-269. https://doi.org/10.5152/jtgga.2012.43

Barragan-Montero, A., Javaid, U., Valdes, G., Nguyen, D., Desbordes, P., Macq, B., . . . Lee, J. A. (2021). Artificial intelligence and machine learning for medical imaging: A technology review. *Phys. Med., 83*, 242-256. https://doi.org/10.1016/j.ejmp.2021.04.016

Boice, E. N., Hernandez-Torres, S. I., & Snider, E. J. (2022). Comparison of Ultrasound Image Classifier Deep Learning Algorithms for Shrapnel Detection. *Journal of Imaging, 8*(5), 140. https://doi.org/10.3390/jimaging8050140

Bulawit, G., Palaoag, T., & Bulawit Jr., B. (2023). Android-based Corn Disease Automated Recognition Tool Using Convolutional Neural Network. *Int. J. Exp. Res. Rev., 30*, 236-246. https://doi.org/10.52756/ijerr.2023.v30.021

Chen. C, & Fajin, D. (2022). Dataset for Fetus Framework", Mendeley Data, V1, https://doi.org/10.17632/n2rbrb9t4f.1.

Cho, H. Y., Kim, Y. H., Park, Y. W., Kim, S. Y., Lee, K. H., Yoo, J. S., & Kwon, J. Y. (2015). Image Settings

Affecting Nuchal Translucency Measurement Using Volume NT Software. *Yonsei Med J, 56*(5), 1345-1351. https://doi.org/10.3349/ymj.2015.56.5.1345

Crossley, J. A., Aitken, D. A., Cameron, A. D., McBride, E., & Connor, J. M. (2002). Combined ultrasound and biochemical screening for Down's syndrome in the first trimester: a Scottish multicentre study. *BJOG, 109*(6), 667-676. https://doi.org/10.1111/j.1471-0528.2002.01394.x

Dominic-Gabriel, I., & Roxana-Cristina, D. (2018). Prenatal Diagnosis of Down Syndrome. In *Advances in Research on Down Syndrome*: InTech.

Driscoll, D. A., Gross, S. J., Professional, P., & Guidelines, C. (2008). First trimester diagnosis and screening for fetal aneuploidy. *Genetics in medicine : Official Journal of the American College of Medical Genetics, 10*(1), 73-75. https://doi.org/10.1097/GIM.0b013e31815efde8

Haloi, R., Chanda, D., Hazarika, J., & Barman, A. (2023). Statistical feature-based EEG signals classification using ANN and SVM classifiers for Parkinson's disease detection. *Int. J. Exp. Res. Rev.*, *31*(Spl), 141-149.

Hill, M., Johnson, J.-A., Langlois, S., Lee, H., Winsor, S., Dineley, B., . . . Chitty, L. S. (2016). Preferences for prenatal tests for Down syndrome: an international comparison of the views of pregnant women and health professionals. *European journal of human genetics : EJHG, 24*(7), 968-975. https://doi.org/10.1038/ejhg.2015.249

Kaur, P. (2023). Performance and Accuracy Enhancement During Skin Disease Detection in Deep Learning. *Int. J. Exp. Res. Rev.*, *35*, 96-108. https://doi.org/10.52756/ijerr.2023.v35spl.009

Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J. Ambient. Intell. Humaniz Comput., 14*(7), 8459-8486. https://doi.org/10.1007/s12652-021-03612-z

Li, L., Liu, W., Zhang, H., Jiang, Y., Hu, X., & Liu, R. (2019). Down Syndrome Prediction Using a Cascaded Machine Learning Framework Designed for Imbalanced and Feature-correlated Data. *IEEE Access, 7*, 97582-97593. https://doi.org/10.1109/access.2019.2929681

Li, L., Ma, C., & Sun, L. (2022). Study on the Effect of B-Ultrasound NT Scan in Early Pregnancy Combined with Serum Screening in Early and Middle Pregnancy for Down Syndrome. *Comput Math Methods Med, 2022*, 7517112. https://doi.org/10.1155/2022/7517112

Lim, C., Inagaki, M., Shinozaki, T., & Fujita, I. (2023). Analysis of convolutional neural networks reveals the computational properties essential for subcortical processing of facial expression. *Scientific reports, 13*(1), 10908-10908. https://doi.org/10.1038/s41598-023-37995-0

Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S. X., . . . Wang, T. (2019). Deep Learning in Medical Ultrasound Analysis: A Review. *Engineering, 5*(2), 261-275. https://doi.org/10.1016/j.eng.2018.11.020

Nia, N. G., Kaplanoglu, E., & Nasab, A. (2023). Evaluation of artificial intelligence techniques in disease diagnosis and prediction. *Discov. Artif Intell, 3*(1).

Niknejadi, M., & Haghighi, H. (2015). Chromosomally and Anatomically Normal Fetuses With Increased First Trimester Nuchal Translucency Conceived by ICSI. *Iran J. Radiol., 12*(2), e7157. https://doi.org/10.5812/iranjradiol.7157

Reddy, N. S., & Khanaa, V. (2023). Diagnosing and categorizing of pulmonary diseases using Deep learning conventional Neural network. *Int. J. Exp. Res. Rev.*, *31*(Spl Volume), 12-22. https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.002

Reshi, A. A., Ashraf, I., Rustam, F., Shahzad, H. F., Mehmood, A., & Choi, G. S. (2021). Diagnosis of vertebral column pathologies using concatenated resampling with machine learning algorithms. *PeerJ. Computer science, 7*, e547-e547. https://doi.org/10.7717/peerj-cs.547

Roozbeh, N., Azizi, M., & Darvish, L. (2017). Pregnancy Outcome of Abnormal Nuchal Translucency: A Systematic Review. *J. Clin. Diagn. Res., 11*(3), QC12-QC16. https://doi.org/10.7860/JCDR/2017/23755.9384

Roozbeh, N., Azizi, M., & Darvish, L. (2017). Pregnancy Outcome of Abnormal Nuchal Translucency: A Systematic Review. *Journal of Clinical and Diagnostic Research : JCDR, 11*(3), QC12-QC16. https://doi.org/10.7860/JCDR/2017/23755.9384

Russo, M. L., & Blakemore, K. J. (2014). A historical and practical review of first trimester aneuploidy screening. *Semin Fetal Neonatal Med., 19*(3), 183-187. https://doi.org/10.1016/j.siny.2013.11.013

Rustam, F., Reshi, A. A., Ashraf, I., Mehmood, A., Ullah, S., Khan, D. M., & Choi, G. S. (2020). Sensor-Based Human Activity Recognition Using Deep Stacked Multilayered Perceptron Model. *IEEE Access, 8*,

218898-218910. https://doi.org/10.1109/access.2020.3041822

Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: a survey. *Evol Intell.*, *15*(1), 1-22. https://doi.org/10.1007/s12065-020-00540-3

Shafi, S., Reshi, A., Shah, A., Al-Mutairi, R. H., Al-Mutairi, G., & Ahmad, A. (2024). Artificial Intelligence Driven Bibliometric Insights: Pioneering Down Syndrome Research. International Journal of Experimental Research and Review, *37*(Special Vo), 68-84. https://doi.org/10.52756/ijerr.2024.v37spl.006

Snider, E. J., Hernandez-Torres, S. I., & Boice, E. N. (2022). An image classification deep-learning algorithm for shrapnel detection from ultrasound images. *Scientific reports*, *12*(1), 8427-8427. https://doi.org/10.1038/s41598-022-12367-2

Sperling, K., Scherb, H., & Neitzel, H. (2023). Population monitoring of trisomy 21: problems and approaches. *Molecular cytogenetics*, *16*(1), 6-6. https://doi.org/10.1186/s13039-023-00637-1

Taye, M. M. (2023). Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions. *Computation*, *11*(3), 52. https://doi.org/10.3390/computation11030052

Van den Veyver, I. B. (2016). Recent advances in prenatal genetic screening and testing. *F1000Res*, *5*, 2591. doi:10.12688/f1000research.9215.1

Wang, Y., Ge, X., Ma, H., Qi, S., Zhang, G., & Yao, Y. (2021). Deep Learning in Medical Ultrasound Image Analysis: A Review. *IEEE Access*, *9*, 54310-54324. https://doi.org/10.1109/access.2021.3071301

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, *9*(4), 611-629. https://doi.org/10.1007/s13244-018-0639-9

Yousef, R., Gupta, G., Yousef, N., & Khari, M. (2022). A holistic overview of deep learning approach in medical imaging. *Multimed Syst*, *28*(3), 881-914. https://doi.org/10.1007/s00530-021-00884-5

Zafar, A., Aamir, M., Nawi, N. M., Arshad, A., Riaz, S., Alruban, A., . . . Almotairi, S. (2022). A Comparison of Pooling Methods for Convolutional Neural Networks. *Applied Sciences-Basel*, *12*(17). https://doi.org/ARTN 8643 10.3390/app12178643

How to cite this Article:

Aijaz Ahmad Reshi, Shabana Shafi, Irfan Qayoom, Maria Wani, Shahida Parveen and Ajaz Ahmad (2024). Deep Learning based architecture for Down syndrome assessment during early pregnancy using fetal ultrasound images. *International Journal of Experimental Research and Review*, *38*, 182-193.

DOI: https://doi.org/10.52756/ijerr.2024.v38.017