Please cite the Published Version

Additional Information: This is an open access article which first appeared in Computers in Biology and Medicine
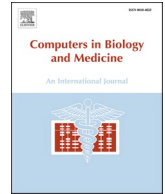
Prediction of fetal brain gestational age using multihead attention with Xception

Mohammad Asif Hasan [a], Fariha Haque [a], Tonmoy Roy [b], Mahedi Islam [a], Md Nahiduzzaman [c], Mohammad Mahedi Hasan [d], Mominul Ahsan [e], Julfikar Haider [f,*]

[a] *Department of Electronics & Telecommunication Engineering, Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh*
[b] *Department of Data Analytics & Information Systems, Utah State University, Old Main Hill, Logan, UT, 84322 (435) 797-1000, USA*
[c] *Department of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Rajshahi, 6204, Bangladesh*
[d] *Department of Apparel Engineering, Textile Engineering College Noakhali, TEC Road, Chowmuhani, Noakhali, 3821, Bangladesh*
[e] *Department of Computer Science, University of York, Deramore Lane, Heslington, York, YO10 5GH, UK*
[f] *Department of Engineering, Manchester Metropolitan University, Chester Street, Manchester M1 5GD, UK*

ABSTRACT

Accurate gestational age (GA) prediction is crucial for monitoring fetal development and ensuring optimal prenatal care. Traditional methods often face challenges in terms of precision and prediction efficiency. In this context, leveraging modern deep learning (DL) techniques is a promising solution. This paper introduces a novel DL approach for GA prediction using fetal brain images obtained via magnetic resonance imaging (MRI), which combines the strength of the Xception pretrained model with a multihead attention (MHA) mechanism. The proposed model was trained on a diverse dataset comprising 52,900 fetal brain images from 741 patients. The images encompass a GA ranging from 19 to 39 weeks. These pretrained models served as feature extraction components during the training process. The extracted features were subsequently used as the inputs of different configurable MHAs, which produced GA predictions in days. The proposed model achieved promising results with 8 attention heads, 32 dimensionality of the key space and 32 dimensionality of the value space, with an R-squared (R^2) value of 96.5 %, a mean absolute error (MAE) of 3.80 days, and a Pearson correlation coefficient (PCC) of 98.50 % for the test set. Additionally, the 5-fold cross-validation results reinforce the model's reliability, with an average R^2 of 95.94 %, an MAE of 3.61 days, and a PCC of 98.02 %. The proposed model excels in different anatomical views, notably the axial and sagittal views. A comparative analysis of multiple planes and a single plane highlights the effectiveness of the proposed model against other state-of-the-art (SOTA) models reported in the literature. The proposed model could help clinicians accurately predict GA.

1. Introduction

Age-related changes in the human brain occur throughout life due to a complex combination of biological and genetic factors [1]. Given that age-related brain alterations are known to be region-specific and impact behavior and cognition, magnetic resonance imaging (MRI) may be a vital biomarker for brain health since age-related alterations in the brain are region-specific and have been linked to behavioral and cognitive functions [2,3]. The brain ages of healthy adults and children have been accurately predicted by structural MRI research in the past [4]. Patients had significantly greater predicted age differences (PADs) than healthy

subjects did, which is thought to be a risk factor for abnormal brain aging or maturation. A large PAD is linked to the severity of the condition, deterioration in cognitive function, and the possibility of developing Alzheimer's disease in the future, according to several long-term studies [5]. As a result, precise brain age estimation can offer medically meaningful insights into brain health, enabling the identification of anatomical anomalies linked to brain illnesses and the prediction of future hazards.

With increasing GA, the human fetal brain exhibits dynamic structural changes. Quantitative structural brain parameters under strict genetic control, such as surface area, curvature, sulcal depth, brain

volume, cortical gyrification, and modest intersubjective differences, are strongly linked with GA in typically developing (TD) fetuses [6]. Furthermore, compared to those in TD fetuses, abnormal structural characteristics have been detected more frequently in fetuses with ventriculomegaly, congenital cardiac disease, and Down syndrome [7]. Thus, fetal brain age prediction using structural MRI features may be helpful in identifying brain development problems and the likelihood of unfavorable developmental consequences.

For several reasons, it is essential to estimate GA accurately in obstetrics. First and foremost, it is crucial for individual women when making obstetric decisions, such as determining which patients would benefit from treatments such as magnesium sulfate for neuroprotection or steroids for fetal lung maturation [8]. It is also crucial when interpreting diagnostic data such as malpresentation or a low-lying placenta, which are only pertinent in the near future. Second, understanding GA is essential at the neonatal level to identify various forms of tiny, vulnerable neonates, such as those who are small for GA or small due to preterm birth, so that they can receive the care they need. Finally, as preterm delivery is the world's leading cause of death for children under five years of age, understanding GA at the population level is crucial to comprehending the prevalence of preterm birth [9]. Although it is feasible to estimate GA at delivery using neonatal assessment, this information is not available for prenatal care, and estimations frequently differ greatly from the gold standard by as much as three to four weeks. The use of biomarkers, such as those in maternal serum and urine, umbilical cord blood, and newborn heel prick testing, is one method to improve the assessment of GA [10]. An accurate, dependable, and reasonably priced biomarker that could estimate GA in this manner would undoubtedly be advantageous. Although a number of potential biomarkers have been studied, no systematic assessment of the ability of these biomarkers to accurately measure GA has been performed.

The research motivation can be summarized as follows:

(a) To enhance the understanding of key aspects of fetal brain maturation, a precise timetable for cortical gyrification and morphological changes was established.
(b) To fill the research gap in fetal brain-based age estimation, prenatal care and outcomes should be improved by enabling early identification of developmental deviations for timely intervention.
(c) To develop a more accurate and robust method for predicting GA from fetal brain MR images, challenges such as variable image quality and motion artifacts must be overcome.

DL presents a promising avenue for GA prediction from fetal brain MR images, offering several advantages over traditional manual systems. First, DL algorithms have the capacity to automatically learn intricate patterns and features from large datasets, thus potentially enhancing the accuracy and robustness of GA predictions. DL-based GA prediction models have the potential to streamline clinical workflows and reduce reliance on manual assessment methods. By automating the GA estimation process, clinicians can save time and resources, allowing for more efficient prenatal screening and monitoring practices.

The aim of this work was to predict GA from fetal brain MR images, an area with very limited existing research, highlighting the novelty and significance of this work. The main contributions of this research are as follows:

1. This research proposes a customized and novel DL model by combining the strengths of Xception with MHA for feature extraction and regression to handle the various spatial perspectives of fetal brain MRI datasets to achieve state-of-the-art GA prediction.
2. Unlike most existing studies that rely on single-plane imaging, the effectiveness of the proposed model was analyzed across different imaging perspectives (axial, coronal, and sagittal): single planes, combinations of any two planes and combinations of all three planes.

3. The performance of the proposed model was compared with that of different transfer learning models (TLMs) through an ablation study to further demonstrate its effectiveness for GA prediction.
4. The performance of the model was also compared with that of SOTA models in terms of GA prediction to demonstrate its competitiveness and potential advancements in the field.

2. Literature review

The importance of accurately predicting GAs has led researchers to employ diverse methodologies for this purpose. Several studies have been conducted using different ML and DL algorithms on MR images or ultrasonogram-based biometrics to predict GA accurately.

Lu et al. [11] reported promising results when an ensemble model consisting of random forest, XGBoost, and LightGBM algorithms was used. This study highlighted that models incorporating both Fetal Fraction (FF) and Routine Clinical Tests (RCTs) yielded more accurate GA predictions than those using only FF or RCTs individually. The ensemble model achieved 64.3 % accuracy, with a mean relative error of 7. They suggested that a hybrid approach that combines multiple sources of data might enhance GA prediction accuracy. Additionally, using advanced modeling techniques or ensembles could further improve predictions, especially for the challenging 19–38-week range that is of particular interest in this case. However, they did not mention whether they predict age in days or weeks. Another study by Diego et al. [12] developed an ML pipeline using multimodal fetal MRI data to predict GA at birth, achieving an R^2 of 0.51 and a mean absolute error of 2.22 weeks. The pipeline demonstrated high accuracy (0.88), sensitivity (0.86), and specificity (0.89), outperforming previous methods. The key predictive features included cervical length and placental T2 values. This work provides a proof of concept for using regression models in preterm birth prediction and suggests future expansion to improve stratification within the preterm cohort. A study by Ross et al. [13] developed a logistic regression model to predict adverse immediate neonatal adaptation (INA) in fetuses with suspected severe fetal growth restriction after 34 gestational weeks. Using data from 1220 women, the model incorporated six prelabor features and achieved an area under the curve of 78 %, with a sensitivity of 66 % and a specificity of 83 %. The strong performance of the model demonstrated the potential to predict INA risk using prelabor characteristics, suggesting early intervention strategies. Another study by Mahmood et al. [14] presented an end-to-end framework for segmenting, measuring, and estimating fetal GA and weight from 2D ultrasound images of the fetal head. Using eight fine-tuned segmentation architectures and a weighted voting method to create an ensemble transfer learning model (ETLM), the framework achieved high segmentation accuracy (98.53 % mean intersection over union), precise measurements (1.87 mm mean absolute difference), and low prediction errors for the GA (0.03 % MSE) and estimated fetal weight (0.05 % MSE). Validation against expert assessments and longitudinal references confirms the framework's reliability and superiority over SOTA methods. Another study presented by Lee et al. [15] presented a SOTA ML model for estimating GA using standard ultrasound images without measurement data. After training and validation on two independent datasets, the model was blinded to the ground truth of the GA. It compensates for variations in fetal size and remains accurate even in cases of intrauterine growth restriction. The model estimates GA with a mean absolute error of 3.0 days in the second trimester and 4.3 days in the third trimester, outperforming current ultrasound-based clinical methods. This approach improves the accuracy of pregnancy dating, particularly later in gestation.

Shen et al. [16] introduced an end-to-end, attention-guided DL model for predicting fetal brain MR images. The model achieved an R^2 score of 0.945, a mean absolute error of 6.7 days, and a concordance correlation coefficient of 0.970. After training on 741 images from 19 to 39 weeks gestation, the model demonstrated strong generalizability across independent datasets from four institutions, achieving R^2 scores

of 0.81–0.90 with minimal fine-tuning. This regression algorithm offers a robust, automated tool for assessing in utero neurodevelopment and estimating GA beyond the first trimester. Liyue Shen et al. [17] implemented an end-to-end model with an attention-guided DL method to predict the GA. The model results included an R^2 value of 0.945, a mean absolute error of 6.7 days, and a concordance correlation coefficient of 0.970. They used a heterogeneous dataset of 741 developmentally normal fetal brain images to build a convolutional neural network. Zhao et al. [18] proposed an attention-based Hemispheric Relation Inference Network (HRINet) to leverage the inherent structural lateralization of the brain during early development. This model captures the inter-hemispheric relationships using a graph attention mechanism, transmitting lateralization information as features to describe the interactive development between the bilateral hemispheres. The HRINet was applied to estimate the brain age of 531 preterm and full-term neonates from the Developing Human Connectome Project (dHCP) database. Compared with other benchmark models, HRINet demonstrated superior performance in predicting perinatal brain age, achieving a mean absolute error of 0.53 and a determination coefficient (R^2) of 0.89. The generalizability of HRINet was further validated on an independent dataset collected from the Gansu Provincial Maternity and Child Care Hospital. Additionally, when applied to a separate dataset consisting of 47 scans of preterm infants at term-equivalent age, the model predicted significantly lower ages than chronological age did, indicating delayed brain development in premature infants. However, this study has several limitations. The sample size was relatively small, and although the model's effectiveness and generalizability were validated on an additional independent dataset, further testing on a larger dataset with diverse demographic groups is necessary to assess the model's generalizability fully. Moreover, while chronological age has been commonly used as the ground truth in brain age prediction studies, it may not fully capture the individual variations in brain development, highlighting the need for more comprehensive metrics in future research. Kojita et al. [19] trained a DL model with T2-weighted images from 126 training cases and 29 validation cases. Although the predicted value of ρc for BPD (ρc = 0.920) was moderate, the model's ρc (ρc = 0.964) was substantial. An attention-based DL model was used to predict a GA with an end-to-end framework by Liyue Shen et al. [20] This work combines key insights from multiview MR images, including axial, coronal, and sagittal MR images, with age-activated weakly supervised attention maps. The extensive experiments revealed an age prediction performance with an R^2 of 0.94 via multiview MRI and attention. Ran et al. [21] conducted a study at Wuhan Children's Hospital, and a total of 1, 327 MR images were collected from 157 healthy fetuses between 22 and 34 weeks of gestation. These images were utilized to evaluate the performance of a novel algorithm, JoCoRank, designed for fetal brain age estimation. The JoCoRank algorithm, which incorporates joint correlation learning with ranking similarity regularization, demonstrated promising results, with an average MAE of 0.693 ± 0.064 weeks and an R^2 coefficient of 0.930 ± 0.019. Despite its success, the study has certain limitations. The method employed by JoCoRank focuses solely on single-view (coronal orientation) fetal brain MR images and does not leverage the information available from other orientations, such as axial and sagittal views. Ziteng et al. [22] introduced the Pyramid Squeeze Attention (PSA)-guided Dynamic Feature Fusion CNN (PDFF-CNN), which is designed to robustly predict gestational age from fetal brain MR images, particularly from an imbalanced dataset. The PDFF-CNN framework consists of four key components: a transformation module, a feature extraction module, a dynamic feature fusion module, and a balanced mean square error (MSE) loss function. The transformation and feature extraction modules utilize the PSA to learn multiscale and multi-orientation feature representations through a parallel weight-sharing Siamese network. The PDFF-CNN was evaluated on an imbalanced dataset comprising 1,327 routine clinical T2-weighted MR images from 157 subjects. The model demonstrated promising performance in predicting gestational age, achieving a mean absolute error of

0.848 weeks and an R^2 of 0.904. Additionally, attention activation maps generated by PDFF-CNN highlighted regional features that significantly contributed to gestational age prediction at various stages of gestation. A limitation of this study is that it only involves the prediction of gestational age for healthy fetuses, which may restrict the generalizability of the model to other populations. Mengting et al. [23] used 577 T1 MR images of preterm neonates from two different datasets. The NEOCIVET pipeline was used to generate cortical surfaces and morphological features, which were then input into a Graph Convolutional Network (GCN) to predict brain age. The brain age index (BAI) is defined as the difference between the predicted brain age (PBA) and chronological age. The results indicated that brain morphology and GCN-based age prediction for preterm neonates without brain lesions (mean absolute error: 0.96 weeks) outperformed conventional machine learning methods that do not utilize topological information. Structural equation modeling (SEM) revealed that BAI mediated the effects of preterm birth and postnatal clinical factors on neurodevelopmental outcomes at 30 months but not the effects of perinatal brain injuries. Despite the strengths of this study, including its large dataset, it has several limitations. The dataset was heterogeneous, consisting of data from two different sources, and the MRI scans from UCSF were acquired using a different protocols and MRI strengths. Although harmonization techniques were applied, the number of participants who underwent clinical evaluations and follow-up neurodevelopmental assessments was relatively small. Additionally, the significant time gap between the neonatal period and infancy at 30 months may limit the model's predictive power. Ansari et al. [24] presented a study in which they used a DL approach with a CNN based on the Inception architecture to estimate brain age from 20-min, single-channel EEG recordings in preterm neonates, achieving a mean absolute error (MAE) of 0.78 weeks. Compared with traditional techniques, this method significantly reduces the recording duration and complexity. The model effectively differentiates between normal and severely abnormal neurodevelopmental outcomes and demonstrates robust accuracy on an independent dataset, highlighting its potential for clinical application. Using routine clinical T2-weighted MR images of 659 fetal brains, Wen et al. [25] developed an attention-based deep residual network for fetal brain age prediction using T2-weighted MR images, achieving a mean absolute error of 0.767 weeks and an R^2 of 0.920. The model quantified predictive uncertainty to detect anomalies such as a small head circumference and malformations with high diagnostic power (AUC of 0.90). Attention maps revealed key regional features for age estimation. This approach shows promise for clinical prenatal diagnosis.

Kim et al. [26] investigated functional brain connectivity development in utero during the second and third trimesters using 110 resting-state functional magnetic resonance imaging (MR) scans from 95 healthy fetuses. Representational similarity analysis revealed that intersubjective similarity in fetal functional connectome patterns correlated with GA differences (r = 0.28, P < 0.01) and showed GA sensitivity, particularly in the frontal area. A critical subnetwork predicted GA with a mean absolute error of 2.72 weeks, and functional connectome patterns reliably predicted individual fetuses' GA (r = 0.51, P < 0.001). The study also revealed a principal brain network that tracks fetal brain maturity, resembling the global signal in the adult brain. Mazher et al. [27] presented an automatic multitissue fetal brain segmentation model using inception residual and dense spatial attention blocks. Three GA prediction methods, including radiomics-based approaches, achieved strong performance (RMSE: 1.42), as validated on 80 fetal brain MR images. The effectiveness of IRMMNET, demonstrated by a Dice score of 0.791, highlights its potential for clinical applications such as tumor segmentation. Yasuyuki et al. [28] evaluated the accuracy of a DL model in predicting GA from fetal brain MR images by comparing it with biparietal diameter (BPD) measurements. The model, trained on 126 patients and validated on 29 patients, demonstrated substantial concordance (ρc = 0.964), outperforming the BPD model (ρc = 0.920). However, both methods show increasing discrepancies from the

reference standard as the GA increases, with the model's upper limit of prediction being significantly shorter than that of the BPD (2.45 weeks vs. 5.62 weeks). DL holds promise for accurate GA estimation from fetal brain MR images acquired after the first trimester. Farzan et al. [29] developed an AI model to predict GA using biometric measurements from fetal brain MR images, specifically BPD, frontooccipital diameter (FOD), and head circumference (HC) measurements. Using a dataset of 52 normal fetal MR images, the AI model's predictions showed strong correlations with manual measurements and high accuracy, particularly with HCs. The results varied depending on the reference used, with Pearson correlation coefficients greater than 0.97 for all comparisons. This AI approach enhances accuracy and convenience in estimating GA, demonstrating significant potential for improving prenatal care through precise fetal development assessment and pregnancy monitoring.

While some authors have attempted to predict GAs, their results have often been subpar, estimating age in weeks and lacking precision. Furthermore, very few studies have focused on predicting GAs between 19 and 38 weeks, as this period is longer than the other periods mentioned in SOTA; therefore, a highly efficient model for handing significant age variation is needed to correctly predict GAs. The development of models that can handle this longer range will be crucial for improving prediction accuracy across a wider gestational spectrum. Furthermore, most studies did not analyze results using both single-plane and multiplane imaging; they typically showcased only one plane. Additionally, many studies have used a limited number of patients as well as a limited number of training images, resulting in models that lack robustness. Addressing these gaps by employing advanced DL techniques, focusing on direct age prediction, utilizing multiplanar imaging, and involving larger patient cohorts will increase the accuracy and reliability of GA prediction models.

3. Methodology

The methodology section outlines the various steps and techniques employed in this research, from the dataset description to the final evaluation of the proposed model's performance. Below, the overall model architecture, dataset preprocessing, and specific methodologies used are discussed.

3.1. Overall model architecture

The overall model architecture shown in Fig. 1 clearly presents the workflow for predicting the GA. It starts with preprocessing fetal brain MR images by resizing and scaling the images. The dataset is split into training (80 %) and testing (20 %) sets. Initial feature extraction is carried out using the TLM, followed by deep feature extraction with the MHA mechanism to capture intricate relationships within the learned features, enhancing the model's discriminative capabilities, and a dense layer is added for the regression task. It aims to improve spatial hierarchies and computational efficiency. The best-performing model was saved and used to predict the GA. The performance of both the single plane (axial, coronal, and sagittal) and the combination of all three planes (multiple planes) was analyzed by the proposed model. This workflow demonstrates a comprehensive approach for predicting GA from fetal brain MR images using an advanced MHA mechanism combined with Xception for feature extraction and regression tasks.

3.2. Dataset description

The dataset used in this study included fetal brain MR images collected from Stanford Lucile Packard Children's Hospital, which is publicly available. A total of 1927 fetal brain MR images taken between 2004 and 2017 were the subject of the investigation. The 1.5 T (T) and 3 T (T) MRI scans were performed utilizing GE Healthcare's Signa HDxt, Signa EXCITE, Optima MR450W, and Discovery MR750W scanners. The MRI data were collected via an 8-channel head coil. A total of 572 MRI scans that revealed cerebral malformations, ventriculomegaly (enlarged brain ventricles), or other acquired or congenital brain abnormalities were removed from the initial pool of MR images by the researchers. In addition, 422 images were removed because they had too much noise or motion artifacts to be properly interpreted. Following these restrictions, the final database consisted of 933 prenatal brain MR images that had been reviewed by pediatric neuroradiologists and determined to be developmentally normal. The interpretations were mostly based on visual cues from the MR images and biometric measurements, such as the biparietal diameter of the brain (the distance between its two sides) and the occipitofrontal diameter of the skull (the distance between its rear and its front). One-shot rapid spin–echo T2-weighted sequences were
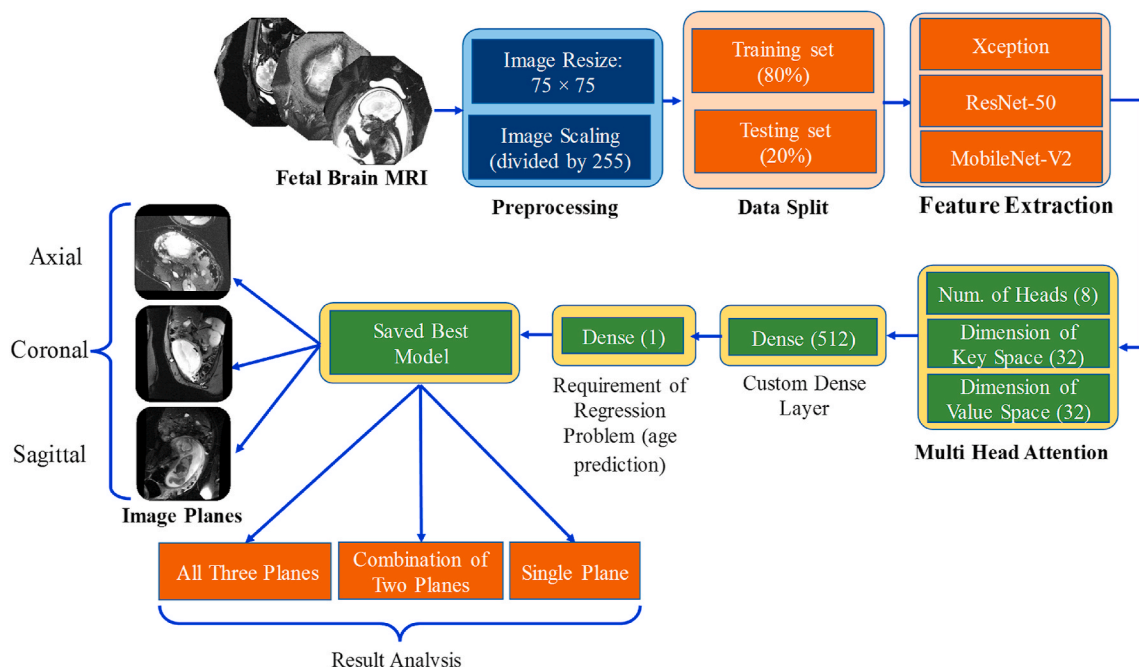

Fig. 1. Proposed workflow for predicting gestational age.

available in all three planes of 741 of the 933 MR images: axial (horizontal), coronal (frontal), and sagittal (side-to-side), as shown in Fig. 2. All the images were then augmented to obtain a total of 52,900 images in the dataset, and the data distribution associated with individual planes is presented in Table 1. For visual display, these single-shot photos that were initially recorded in the Digital Imaging and Communications in Medicine (DICOM) File Format were compressed to JPEG files.

3.3. Dataset preprocessing

The dataset contains 741 individual patient IDs, where 52900 images are stored in black and white jpg formats. All the images were 512×512 pixels long and were too large to be employed in this research environment. Therefore, all the images were resized to 75×75 pixels without degrading the image quality. Next, the images are converted from grayscale to RGB color space to align with the requirements of the proposed model, which requires RGB input. The amount of additional storage space and processing power required is reduced by this downsizing strategy. Images usually use many intensity levels to be represented. After the pixel values were divided by 255, normalization was used to simplify the images and to change the scale from 0 to 255 to 0–1. To ensure the integrity of the study and prevent any bias in the model's predictive performance, the data split was performed on a subject-wise basis rather than on individual images. More specifically, for each subject, 80 % of their images were allocated to the training set, and the remaining 20 % were reserved for the test set. This approach guarantees that no images from the same subject appear in both the training and test sets. This allocation ensured a comprehensive representation across the dataset for effective model training and testing.

3.4. Proposed model description

In this section, an in-depth description of the proposed model is provided, highlighting its unique architectural features and the specific design choices made. Every component that contributes to the model's efficiency is discussed here in detail.

3.4.1. Depthwise separable convolution operation for Xception

In the proposed model, depthwise separable convolutions are used in the deep convolutional neural network architecture commonly referred to as Xception. Classical convolutions have alternatives called depthwise separable convolutions, which are apparently significantly faster to compute. Depthwise separable convolutions, which are a key component of Xception, are designed to decouple the spatial and cross-channel filtering operations typically performed in standard convolutions. Instead of applying individual filters to each input channel, the depthwise separable convolution approach first performs depthwise convolution followed by pointwise convolution. It is designed to reduce the computational complexity of standard convolutions while still capturing meaningful spatial information from the input data.

Table 1
Image distribution of each plane in the fetal brain MRI dataset.

Position	No. of Images	No. of Training Images	No. of Testing Images
Axial	16,811	13,448	3363
Coronal	19,619	15,695	3924
Sagittal	16,470	13,176	3294

3.4.2. Depthwise convolution

In the depthwise convolution stage, each input channel is convolved independently with its corresponding depthwise filter. Depthwise convolution performs spatial filtering within each channel, capturing local spatial features while maintaining the number of channels. This step significantly reduces the computational cost compared with traditional convolutions, where a separate filter is used for each input channel. After the depthwise convolution, the pointwise convolution stage is applied. A pointwise convolution is essentially a 1×1 convolution that operates across channels. It applies a linear combination of 1×1 filters to the output of the depthwise convolution. This step helps capture cross-channel correlations and allows the model to learn complex representations by combining spatially filtered channels.

By separating the spatial and cross-channel filtering operations, depthwise separable convolutions reduce the number of parameters and operations required in comparison to standard convolutions. This reduction leads to improved computational efficiency and enables the training of larger and deeper networks with limited computational resources.

Depthwise convolution generates matching feature maps by applying a single convolution kernel to each input channel. Fig. 3 shows the depthwise and pointwise convolutions of both processes. Here, we assume that the typical convolution kernel K is (D_K, D_K, M, N), the input feature map F is (D_F, D_F, M), the output feature map G is (D_G, D_G, N), and M and N represent the numbers of input and output channels, respectively. The feature map's size is denoted by D. K is divided into two convolutions: a pointwise convolution $(1, 1, M, N)$ and a depthwise convolution $(D_K, D_K, 1, M)$.

The formula of the standard convolution is [30]:

$$G_{k,l,n} = \sum_{i,j,m} \widehat{K}_{i,j,m,n} \times F_{k+i-1,l+j-1,m} \tag{1}$$

The cost of computation for the standard convolution is [30]:

$$D_k \times D_k \times M \times N \times D_F \times D_F \tag{2}$$

The expression for the depthwise separable convolution is [30]:

$$\widehat{G}_{k,l,m} = \sum_{i,j} \widehat{K}_{i,j,m,n} \times F_{k+i-1,l+j-1,m} \tag{3}$$

The computational cost for the depthwise separable convolution is [30]:

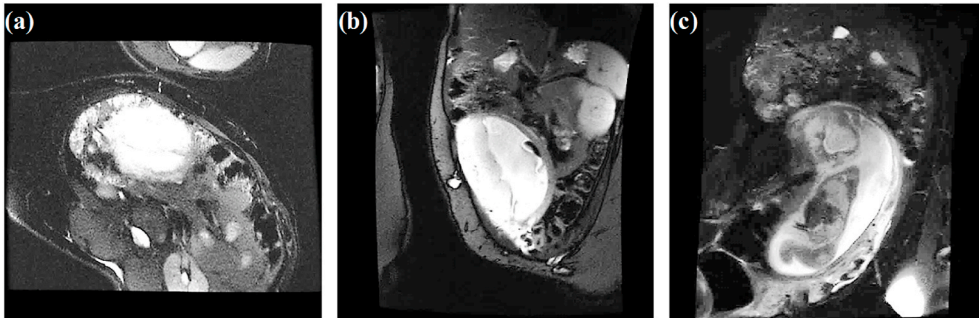$$D_k \times D_k \times M \times D_F \times D_F + M \times N \times D_F \times D_F \tag{4}$$

Fig. 2. Multiple planes of patient ID 1: (a) axial, (b) coronal (c) and sagittal planes.
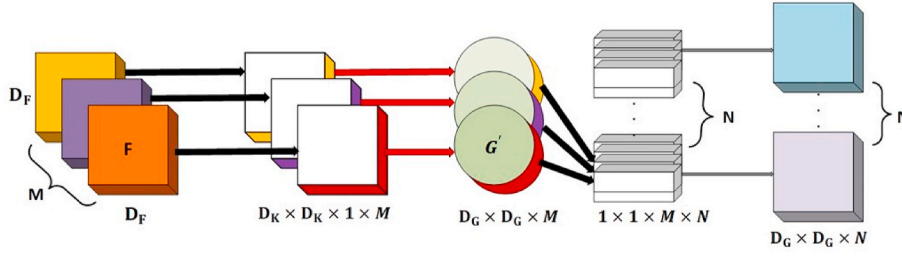
Fig. 3. Block diagram of the depthwise and pointwise convolution operations.

The ratio of calculation consumption (RCC) is defined to determine the computational efficiency of the proposed model [30]:

$$RCC = \frac{D_k \times D_k \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_k \times D_k \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2} \qquad (5)$$

3.4.3. Pointwise convolution

Pointwise convolution, often referred to as a 1×1 convolution, is a type of convolutional operation in which a filter of size 1×1 is applied to each element of the input tensor. Unlike traditional convolutions, which span both spatial dimensions and input channels, pointwise convolutions act only in the channel dimension. It uses filters with a spatial size of 1×1. The term "pointwise" is derived from the fact that the convolutional filter essentially operates at a single spatial point. Then, it performs a convolution operation independently at each spatial location across all the input channels. This means that each element in the output corresponds to a linear combination of the input values at the corresponding spatial location across all channels. One of the primary motivations for using pointwise convolution is parameter reduction. By applying a 1×1 filter independently to each channel, the number of parameters in the network can be significantly reduced, especially compared with standard convolutions that use larger filters. While each pointwise convolution operation is applied independently at each spatial location, the combination of the results across channels effectively allows for the mixing and combination of features. This enables the network to learn nonlinear relationships between different channels, contributing to the expressive power of the model.

3.4.4. Xception architecture

The Xception architecture takes advantage of depthwise separable convolutions throughout its network. It replaces the traditional convolutional layers with depthwise separable convolutional layers. This change allows the Xception model to achieve a high level of accuracy while significantly reducing the number of parameters compared with
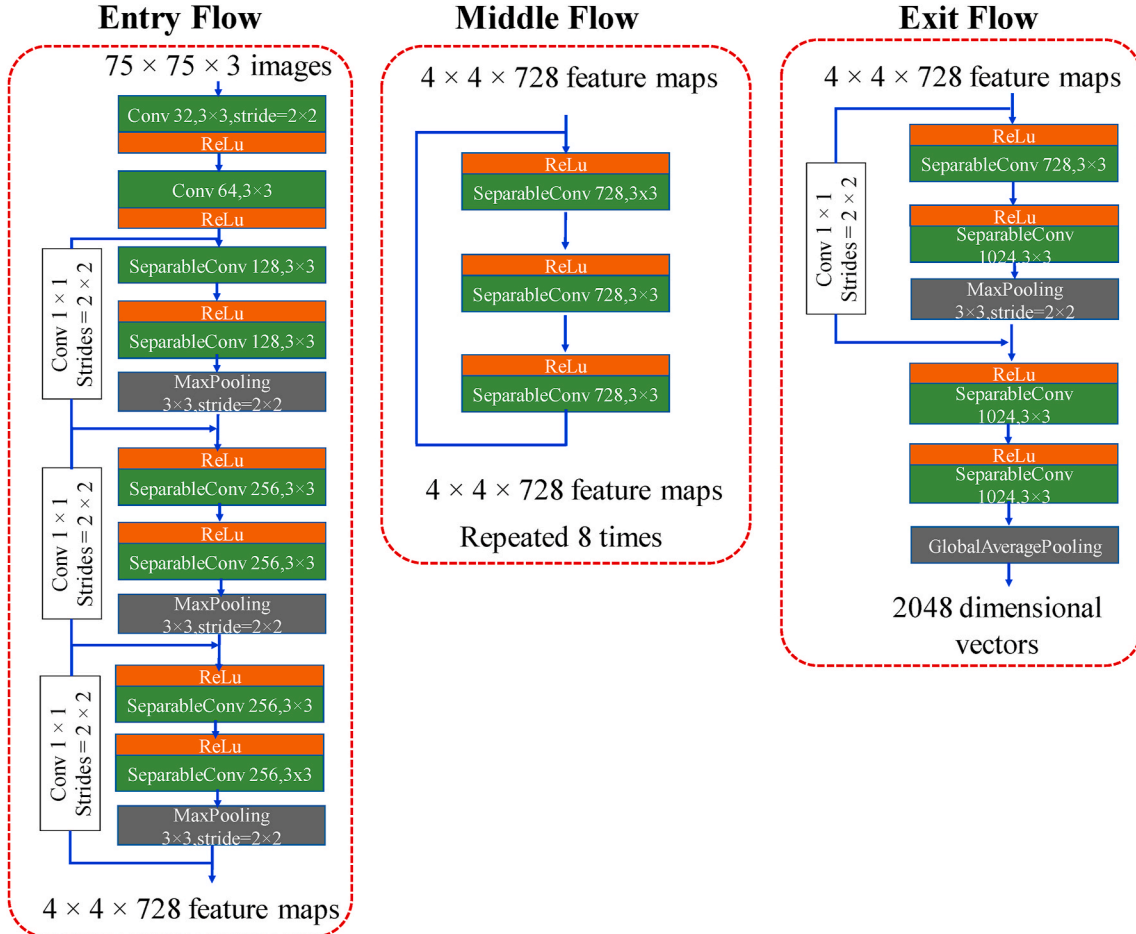

Fig. 4. Each flow input and output dimension for the 75×75 image in the Xception architecture.

earlier CNN architectures. The Xception architecture can be described as a linear sequence of depthwise separable convolution layers, each of which is connected residually. The term "Xception" stands for "Extreme Inception," reflecting its design as a more advanced and powerful iteration of the Inception architecture [31,32].

The input to the network is a 75 × 75 RGB image. The Xception architecture consists of three main sections, which are shown in Fig. 4: the entry flow, middle flow, and exit flow. In the entry flow, three successive blocks, namely, entry block 1, entry block 2, and entry block 3, sequentially apply convolutions and separable convolutions with batch normalization and ReLU activation functions. Each block incorporates a residual connection, and max pooling layers are strategically employed to reduce spatial dimensions. The middle flow comprises repeated blocks of separable convolutions with 3 × 3 kernels and 728 filters, along with batch normalization and ReLU activation functions. Residual connections are applied within each block to facilitate information flow. The exit flow consists of two exit blocks: exit block 1 and exit block 2. Exit block 1 applies separable convolutions with 728 and 1024 filters, followed by batch normalization and rectified linear unit (ReLU) activation. A residual connection and max pooling layer further reduce the spatial dimensions. Exit Block 2 employs separable convolutions with 1536 and 2048 filters, followed by similar normalization and activation functions. The architecture concludes with global average pooling to reduce spatial dimensions. In addition to this architecture, a fully connected (dense) layer and a Softmax activation function were added to generate class probabilities in the final output layer. The dense layer combines high-level features learned by preceding layers, introducing nonlinearity through the Softmax activation function. It is used for parameterized mapping, adjusting weights during training to adapt to data characteristics. This layer aids in dimensionality reduction by condensing complex representations into a suitable format for final predictions. Connected to all the neurons in the previous layer, it acts as a critical element in the neural network's ability to capture intricate relationships within the data. The process of calculating attention in MHA is shown in Fig. 4.

3.4.5. Multihead attention

A module for attention mechanisms known as "multihead attention" operates iteratively and concurrently through an attention mechanism, as depicted in Fig. 5. The main idea behind MHA is to split the input into multiple subspaces and apply attention to each subspace independently.
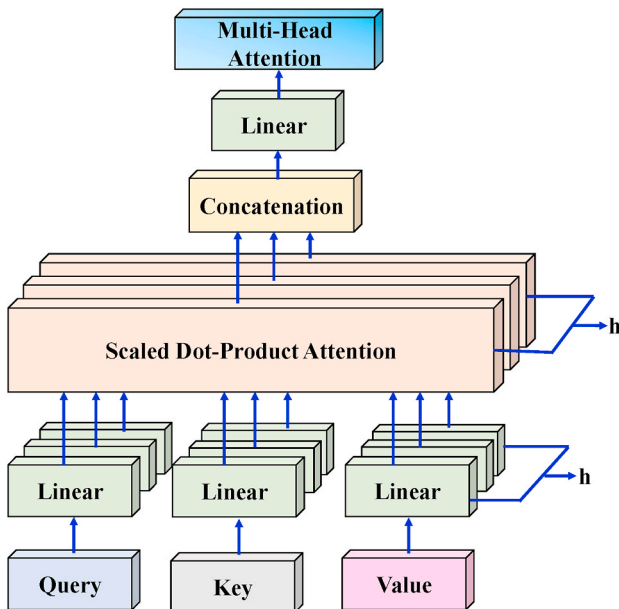

Fig. 5. Mechanism of multihead attention.

Each subspace, also referred to as a head, is associated with its own set of learnable parameters. This allows the model to attend to different parts of the input sequence in parallel, providing a more comprehensive understanding of the data. For collecting useful information from an image, the entire portion is not essential, so the attention mechanism concentrates on the useful portion.

With the use of linear layers and other learnable weights, the input is projected to queries, keys, and values [33].

$$Q = \left(X' \in R^{N \times E}\right) \times \left(W_Q \in R^{E \times d}\right) \tag{6}$$

$$K = \left(X' \in R^{N \times E}\right) \times \left(W_K \in R^{E \times d}\right) \tag{7}$$

$$V = \left(X' \in R^{N \times E}\right) \times \left(W_V \in R^{E \times d}\right) \tag{8}$$

Now, queries (Q), keys (K), and values (V) are obtained. When a scaled dot-product form of attention is used, the computation of the dot product between the queries and keys is scaled by $\sqrt{d_K}$. The attention weights are then obtained by applying the nonlinear Softmax function [33].

$$Z' = \left(Q \in R^{N \times d}\right) \times \left(K^T \in R^{d \times N}\right) \tag{9}$$

$$Z = soft \max\left(\left(Z' \in R^{N \times N}\right) \Big/ \sqrt{d_K}\right) \tag{10}$$

As a result, the scaled dot-product attention function may be expressed in the following shorter form [33]:

$$Attention(Q, K, V) = soft \max\left(\frac{QK^T}{\sqrt{d_K}}\right) V \tag{11}$$

With d-dimensional queries, keys, and values, it is desirable to linearly project them to d_h, d_k, and d_v dimensions, h times, using various learnable weights with linear layers, as opposed to conducting a single attention function. The scaled dot-product attention function is then simultaneously applied to each of the h heads, producing the h number of d_v dimensional values. These attention-weighted values are subsequently combined and concatenated with linear layers [33].

$$Multihead(Q, K, V) = concatinate[head_1, head_2.....head_i]W_o \tag{12}$$

Here, $W_o \in R^{hd \times d}$.

In this research, the use of 8 heads, a key dimension of 32, and a value dimension of 32 in the MHA mechanism proved to be optimal for the dataset. The use of multiple attention heads allows the model to focus on different aspects of the input data simultaneously. This can enhance the model's ability to capture diverse and complex patterns in the dataset. The combination of Xception, known for its depthwise separable convolutions, with MHA provides an increased capacity for the model to learn hierarchical features and relationships within the data. The specified dimensions for the key and value in the attention mechanism are crucial parameters. The choice of 32 for both the key and value dimensions, on the basis of the optimal performance in the study selected by the trial and error approach, indicates adaptability to capturing intricate relationships within the dataset. Overall, the optimal configuration of 8 heads, with key and value dimensions set to 32, in the MHA mechanism, along with the integration of Xception, enables our model to effectively capture diverse patterns, learn hierarchical features, and adapt to intricate relationships, leading to accurate prediction of fetal brain GA.

3.4.6. Dense layers

In the proposed model, a dense layer was incorporated with 512 neurons and a rectified linear unit (ReLU) activation function, which was applied to the output of the attention mechanism. This dense layer plays a crucial role in capturing complex relationships and features within the data, contributing to the model's ability to learn and generalize. The network subsequently concludes with an output layer,

represented by a dense layer, containing 1 neuron designed for regression. This output layer is responsible for producing the final prediction for the GA, encapsulating the collective insights and patterns learned throughout the network. The omission of an activation function in the output layer is deliberate, as this architecture is tailored for regression tasks where the model is expected to directly output a numerical value. The configuration of these layers is a key aspect of the model architecture, representing the transition from learned features to the final prediction, and underscores the model's adaptability for regression tasks.

3.5. System configuration and performance metrics

Table 2 shows the system configuration used to complete the entire computational process.

The necessary fine-tuned hyperparameters used to construct the model are presented in Table 3. Without the correct values of the hyperparameters, the expected results might not be achieved.

For predicting GA in fetal brain MR images, this study focused on the problem of different DL models with MHA, and by identifying the best result among these regressors, the next step was chosen. The best regression model was chosen on the basis of R^2, *MAE, MSE* and *PCC*. R^2 quantifies the extent to which the variability in the dependent variable can be explained by the independent variables. Its value lies between 0 and 1, where higher values signify a superior model fit.

$$R^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\left(\sqrt{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right)\left(\sum_{i=1}^{n}(y_i - \overline{y})^2\right)}\right)} \tag{13}$$

The MAE represents the mean of the absolute differences between the predicted and actual values. It is less affected by outliers than MSE is.

$$MAE = \frac{\sum_{i=1}^{n}|x_i - y_i|}{n} \tag{14}$$

The MSE measures the average squared difference between the predicted values and the actual values. It provides a comprehensive assessment of overall model performance. It provides a more interpretable measure of the average prediction error.

$$MSE = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n} \tag{15}$$

$$PCC = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(\widehat{y}_i - \overline{\widehat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2 \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}} \tag{16}$$

The PCC is a measure of the linear correlation between two variables, providing insight into the strength and direction of the relationship, where y_i and y_i are the actual and predicted values, respectively; $\overline{y}$ is the mean of the actual values; and $\overline{y}$ is the mean of the predicted values. In the comprehensive evaluation of the performance parameters, several

pretrained models with MHA were utilized as regressors, and their effectiveness was assessed using key metrics such as R^2, MSE, and PCC. The performance of each model is presented in the following tables, which highlight their performance metrics.

4. Experimental results and discussion

This section presents the experimental results obtained from the implementation of the proposed model. A detailed analysis of their performance metrics and visual representation is provided here.

4.1. Ablation study with transfer learning models

A comprehensive analysis of the experimental results was obtained through the ablation study presented in this section. Many types of TLMs with custom dense layers and different MHA configurations were tested to construct the proposed model. Table 4 presents the results of the multiplane approach for predicting the GA using the TLM-Dense (1), TLM-Dense (512)-Dense (1) and TLM-MHA-Dense (512)-Dense (1) models on the test set. The Xception-MHA (8-32-32)-Dense (512)-Dense (1) model achieved the highest R^2 of 96.5 %, indicating the best fit among all nine models. It also had the lowest MAE of 3.80 days, suggesting that the model made the most accurate GA predictions. The PCC value of 98.5 % further confirmed the strong correlation between the predicted and actual GA values. The ResNet50-MHA (8-32-32)-Dense (512)-Dense (1) model also performed well, but its accuracy was slightly lower than that of the Xception-MHA (8-32-32)-Dense (512)-Dense (1) model.

Notably, the models constructed with custom dense layers without MHA, such as Xception-Dense (512)-Dense (1), MobileNetV2-Dense (512)-Dense (1), and Xception-Dense (512)-Dense (1), demonstrated better performance than the model without a custom dense layer. The custom dense layer enhances the model's capacity to learn intricate patterns, contributing to improved accuracy. The inclusion of MHA in the proposed model enhanced its ability to capture long-range dependencies and intricate patterns within the data, leading to superior

Table 2
System configurations for developing the computational model.

System configuration	Values
Programming Language	Python
Computational environment	Google colab
Backend	Keras with TensorFlow
Required disk space	78.2 GB
GPU RAM	15 GB
Graphics processing unit (GPU)	Nvidia Tesla T4
GPU random access memory (RAM)	15 GB
System RAM	12.72 GB
Operating system	Windows 11
Input data	MRI Images
Input image size	75 × 75

Table 3
Hyperparameters of the proposed model.

Parameters	Values
Loss function	Mean squared error
Introductory learning rate	0.001
Number of epochs	50
Batch size	16
Shuffling	Every epoch
Optimizer	Adam
Number of attention heads	8
Dimensionality of the key space	32
Dimensionality of the value space	32

Table 4
Multiplane results produced using the TLM-Dense (1), TLM-Dense (512)-Dense (1) and TLM-MHA-Dense (512)-Dense (1) models for the test set.

Model	R-squared	MAE	MSE	PCC
ResNet-50-Dense (1)	90.7	7.43	10.96	93.2
MobileNetV2-Dense (1)	89.7	8.72	120.85	93.7
Xception-Dense (1)	91.3	6.98	95.23	94.3
Xception-Dense (512)-Dense (1)	94.9	4.03	70.78	97.5
MobileNetV2-Dense (512)-Dense (1)	93.6	5.55	88.76	96.8
ResNet50-Dense (512)-Dense (1)	93.5	5.51	90.58	97.1
ResNet50- MHA (8-32-32)-Dense (512)-Dense (1)	95.1	3.82	68.16	97.5
MobileNetV2- MHA (8-32-32)-Dense (512)-Dense (1)	86.7	8.13	185.31	94
Xception-MHA (8–32–32)-Dense (512)-Dense (1)	**96.5**	**3.80**	**48.92**	**98.5**

performance. The absence of MHA in the evaluated models contributed to their comparatively lower performance metrics. Therefore, integrating MHA and custom dense layers into TL models can significantly improve the accuracy and reliability of GA predictions. Among all the combinations, the Xception-MHA (8-32-32)-Dense (512)-Dense (1) model produced the best results.

4.2. K-fold cross-validation

K-fold cross-validation is a robust technique used to assess the generalizability and reliability of a predictive model. This approach reduces the risk of overfitting and provides a more comprehensive understanding of the model's capabilities. Table 5 presents the results of a 5-fold cross-validation (CV) for the proposed Xception-MHA (8-32-32)-Dense (512)-Dense (1) model for predicting the GA. The performance metrics of the proposed model are reported for each fold of the CV, with average values across all folds. The model shows consistently high performance across the folds, with an average R-squared of 95.94 %, indicating that the model explains nearly 96 % of the variance in the data. The average MAE and MSE values are slightly lower, with averages of 3.61 and 58.27, respectively, demonstrating the model's accuracy in predicting the target variable. The high average PCC value of 98.02 % suggests a strong linear relationship between the predicted and actual values. An analysis of the performance across all the folds clearly revealed that the model was not overfit. Overall, the model appeared to perform reliably across different subsets of the data, confirming the generalizability of the proposed model. Fig. 6 shows the regression plots for all 5 folds of the proposed model.

4.3. Results analysis across different MR imaging planes

Notably, the Xception-MHA (8-32-32)-Dense (512)-Dense (1) model outperformed all the other models, demonstrating superior performance across the key metrics. To further understand the robustness of this architecture, several configurations of the MHA component were tested by exploring variations in the number of heads, key dimensions, and value dimensions. These variations provided valuable insights into how different attention mechanisms influence the model's performance, reaffirming that the chosen configuration MHA (8–32–32) produces the highest accuracy and lowest MAE. In addition, the analysis focused on evaluating the effectiveness of different image planes, such as a single plane, a combination of two planes, and three planes, for predicting GA from fetal brain MR images.

4.3.1. Combination of all three planes

In real-life clinical scenarios, the combination of three planes is more suitable because it provides an in-depth view of the fetal brain. Each plane offers unique information: the axial plane can highlight the lateral and third ventricles, the coronal plane can show the brain's symmetry and midline structures, and the sagittal plane can display the corpus callosum and brainstem [34]. Table 6 shows the results of the experiments performed to predict fetal brain GA in 741 patients by taking all three available axial, coronal, and sagittal plane images. Notably, the combination of 8-32-32 (number of attention heads-dimensionality of the key space-dimensionality of the value space) yielded a commendable

R^2 value of 96.5 %, indicating strong predictive performance. This model configuration is characterized by low errors, including an MAE of 3.80 days and a PCC of 98.5 %. Fig. 7 shows the regression plots for the actual and predicted GAs. The choice of hyperparameters, such as the number of attention heads, dimensionality of the key space and dimensionality of the value space, profoundly influences the model's ability to capture intricate relationships within fetal brain MRI data. In this context, the superior performance of the 8-32-32 configuration can be attributed to the increased complexity and expressiveness introduced by having eight attention heads. This enables the model to address different aspects of the input data simultaneously, extracting more nuanced features and patterns related to the GA. Furthermore, the dimensionality of the key space also plays a critical role. A higher dimensionality, as seen in the 8-32-32 configuration, provides the model with a richer representation of key features, facilitating a more accurate prediction of gestation age. However, striking a balance is crucial, as an overly complex model may lead to diminished performance due to overfitting or excessive computational demands.

4.3.2. Combination of two planes

Table 7 presents the performance results of the proposed model when two-plane combinations are used on the test set. The axial-sagittal combination yields the highest performance among the two-plane models, with an R^2 of 95.7 %, an MAE of 4.97 days, and a PCC of 98.4 %. The axial-coronal combination had an R^2 of 90.6 %, an MAE of 5.11 days, and a PCC of 95.4 %. The sagittal-coronal combination performed the least effectively, with an R^2 of 88.2 %, an MAE of 7.52 days, and a PCC of 94.0 %. Compared with the multiplane results (R^2 of 96.5 %, MAE of 3.80 days, and PCC of 98.5 %), the multiplane approach outperforms all two-plane combinations, demonstrating the advantage of utilizing all three planes for more accurate predictions. Compared with the single-plane results, the axial-sagittal combination performed better than the coronal and sagittal planes alone but slightly worse than the axial plane in terms of R^2 and the MAE.

Fig. 8 shows the comparison between the actual and predicted GAs derived from combined fetal brain MR images across two different anatomical planes: axial-sagittal, axial-coronal, and sagittal-coronal. Fig. 8 (a) shows the relationship between the actual GA and the predicted GA when the MR images were taken from the combined axial and sagittal planes. The plot demonstrates a high degree of accuracy (77 %) in the GA predictions, as indicated by the close alignment of the predicted values with the actual values. According to the results, the axial-sagittal combination performed the best, with an R^2 value of 95.7 %, an MAE of 4.97 days, and a PCC of 98.4 %. This high performance can be attributed to the complementary information provided by the axial and sagittal planes, enhancing the model's ability to capture relevant features for accurate GA prediction. Fig. 8 (b) The actual and predicted GAs are plotted for MR images taken from the combined axial and coronal planes. The data points indicate that the model's predictions are reasonably accurate, although not as precise as those of the axial–sagittal combination. The axial-coronal combination had an R^2 value of 90.6 %, an MAE of 5.11 days, and a PCC of 95.4 %. The axial and coronal planes together provide substantial information, but the increased variability in the coronal images likely affects the overall accuracy. Fig. 8 (c) presents the actual versus predicted GA for MR images taken from the combined sagittal and coronal planes. The plot shows that the predicted GA values have more variance than the actual GA values do, indicating the least accurate predictions among the three combinations. The sagittal-coronal combination performed the worst, with an R^2 value of 88.2 %, an MAE of 7.52 days, and a PCC of 94.0 %. The lower performance may be due to the greater variability and less distinct features in the coronal plane, which, when combined with sagittal plane images, do not provide as many robust predictive features as the axial plane does. Overall, Fig. 8 demonstrates the effectiveness of the proposed model in predicting GA from combined fetal brain MR images, with the axial-sagittal combination providing the highest

Table 5
5-Fold CV of the proposed Xception-MHA (8–32–32)-Dense (512)-Dense (1) model.

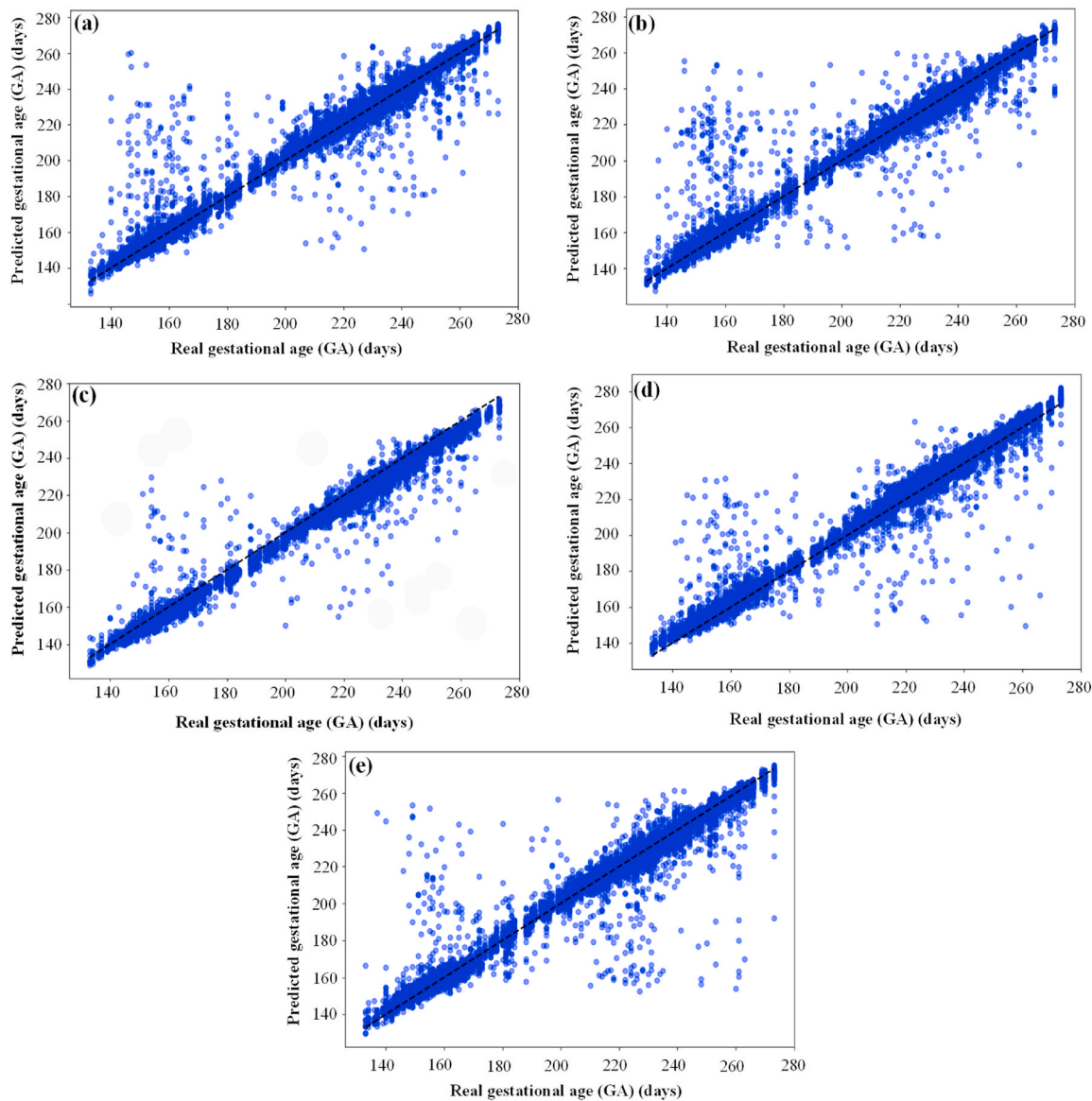| Model	R-squared	MAE	MSE	PCC
Fold-1	95.6	3.51	60.72	97.8
Fold-2	95.8	3.37	67.13	97.6
Fold-3	96.5	3.80	48.92	98.5
Fold-4	96.4	3.94	50.13	98.5
Fold-5	95.4	3.43	64.45	97.7
Average	95.94	3.61	58.27	98.02

Fig. 6. Actual and predicted GA plot of the proposed model Xception-MHA (8-32-32)-Dense (512)-Dense (1) using multiple planes (axial, coronal, and sagittal) of 5-fold CV.

Table 6

Proposed Xception-MHA-Dense (512)–Dense (1) model results obtained via multiple planes by varying the configuration of the multihead attention mechanism.

Model with varying multihead attention	R-squared	MAE	MSE	PCC
8-64-64	96.1	4.12	54.68	98.3
8–32–32	**96.5**	**3.80**	**48.92**	**98.5**
4-64-64	94.8	4.08	71.85	97.5
4-32-32	96.00	3.80	55.95	98.2
2-64-64	95.3	3.69	65.57	97.60
2-32-32	95.60	3.86	60.67	97.90

accuracy, followed by the axial-coronal and then the sagittal-coronal combination. The model's performance is best when the axial plane is used in combination with either the sagittal or the coronal plane because the axial plane has clearer and more consistent features.

4.3.3. Single-plane

The single-plane approach is still valuable in scenarios where

computational resources are limited or when quick assessments are needed. Table 8 presents the performance results of the proposed model when it is applied to single-plane images in the test set. The axial plane yields the highest performance, with an R^2 of 97.1 %, an MAE of 3.24 days, and a PCC of 98.6 %. The sagittal plane had an R^2 of 93.6 %, an MAE of 4.49 days, and a PCC of 96.7 %. The coronal plane performs the least effectively, with an R^2 of 86.3 %, an MAE of 7.06 days, and a PCC of 93.0 %. Compared with the multiplane results of the 8-32-32 configuration, which achieved an R^2 of 96.5 %, the MAE was 3.80 days, and the PCC was 98.5 %. From the above analysis, it was evident that the multiplane (all three-planes) approach leverages the complementary information from all three planes to deliver a more robust and accurate prediction than any single-plane (coronal or sagittal) approach did. The axial plane, however, showed a slightly higher R^2 and lower errors than the three-plane result did, indicating its strong standalone predictive power.

Fig. 9 shows the comparison between the actual and predicted GAs derived from fetal brain MR images across three different anatomical planes: axial, coronal, and sagittal. Fig. 9 (a) shows the relationship between the actual GA and the predicted GA when the MR images were
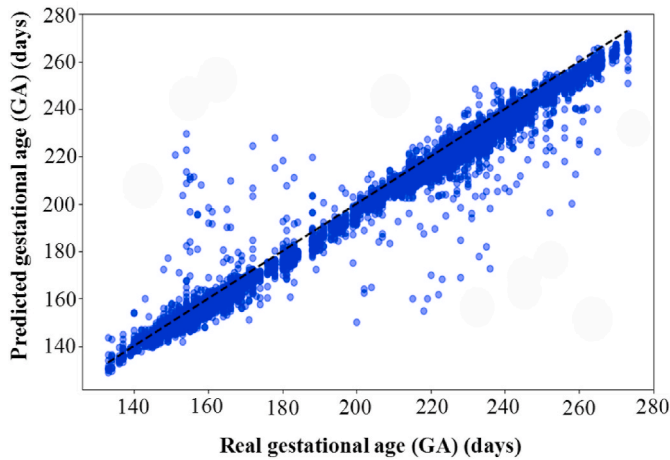
Fig. 7. Actual and predicted GA plot of the proposed model Xception-MHA (8–32–32)-Dense (512)-Dense (1) using multiple planes (axial, coronal, and sagittal).

Table 7
Results of the proposed Xception-MHA (8-32-32)-Dense (512)-Dense (1) model using two planes.

Model	R-squared	MAE	MSE	PCC
Xception-MHA (8-32-32)-Dense (512)-Dense (1) -Axial- Sagittal	95.7	4.97	60.95	98.4
Xception-MHA (8-32-32)-Dense (512)-Dense (1) -Axial- Coronal	90.6	5.11	130.84	95.4
Xception-MHA (8-32-32)-Dense (512)-Dense (1) -Sagittal- Coronal	88.2	7.52	163.40	94.0

taken from the axial plane. The plot demonstrates the accuracy of the GA predictions by aligning the predicted values closely with the actual values, suggesting a high degree of correlation between the two. According to the results, the axial plane performed the best, with an R^2 value of 97.1, an MAE of 3.24 days, and a PCC of 98.6. This high performance can be attributed to the model's ability to capture more distinct and relevant features in axial plane images, which are often clearer and more consistent. The axial-coronal and sagittal-coronal combinations, however, did not perform as well as the single-plane axial model did, highlighting the superior standalone predictive power of the axial plane.

Fig. 9 (b) The actual and predicted GAs are plotted for MR images taken from the coronal plane. The data points indicated that the model's predictions were less consistent with the actual GA values than those in the axial plane were, reflecting the reduced robustness of the model in predicting the GA from coronal MR images. The coronal plane performed the worst, with an R^2 value of 86.3, an MAE of 7.06 days, and a PCC of 93.0. This lower performance may be because the coronal plane images have more variability and potentially fewer distinctive features for the model to accurately predict the GA. Fig. 9 (c) presents the actual versus predicted GA for MR images taken from the sagittal plane. The plot shows that the predicted GA values are relatively well aligned with

Table 8
Results of the proposed Xception-MHA (8-32-32)-Dense (512)-Dense (1) model using a single plane.

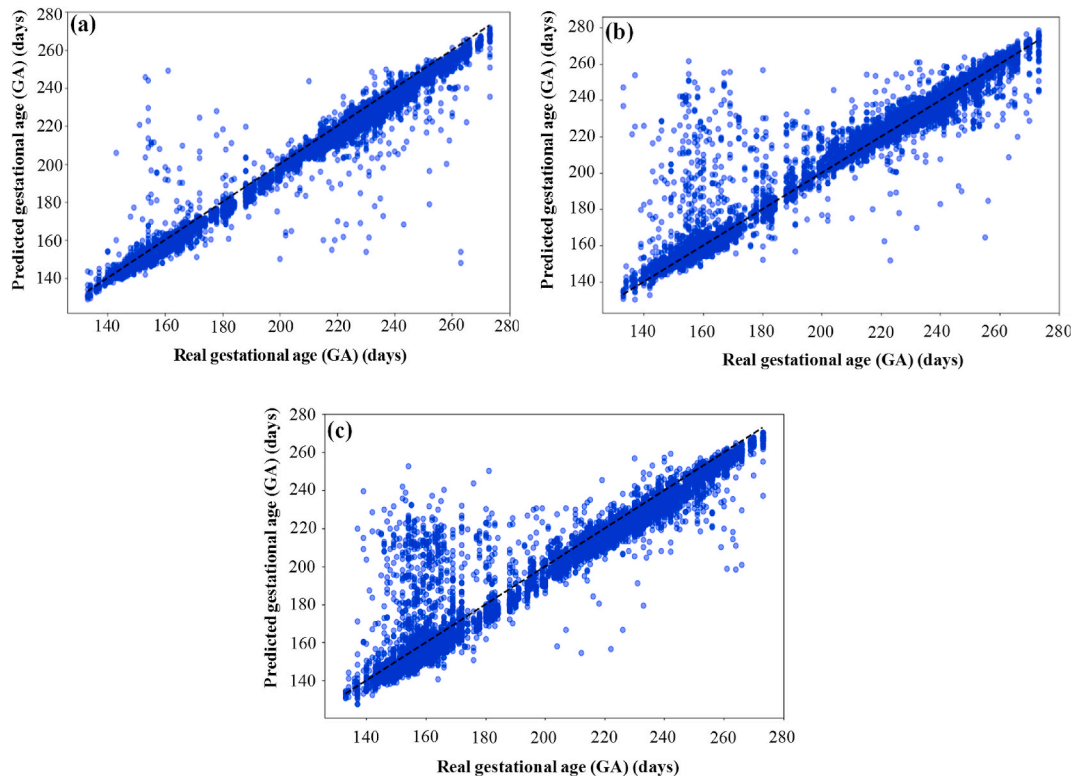| Model	R-squared	MAE	MSE	PCC
Xception-MHA (8-32-32)-Dense (512)-Dense (1) -Axial	97.1	3.24	41.179	98.6
Xception-MHA (8-32-32)-Dense (512)-Dense (1) -Coronal	86.3	7.06	188.22	93.0
Xception-MHA (8-32-32)-Dense (512)-Dense (1) -Sagittal	93.6	4.49	90.32	96.7

Fig. 8. Actual and predicted GA plots of combined two-axis images using two planes for the proposed model Xception-MHA (8-32-32)-Dense (512)-Dense (1) (a) axial-sagittal, (b) axial-coronal and (c) sagittal-coronal images for predicting gestational age in the fetal brain MRI dataset.
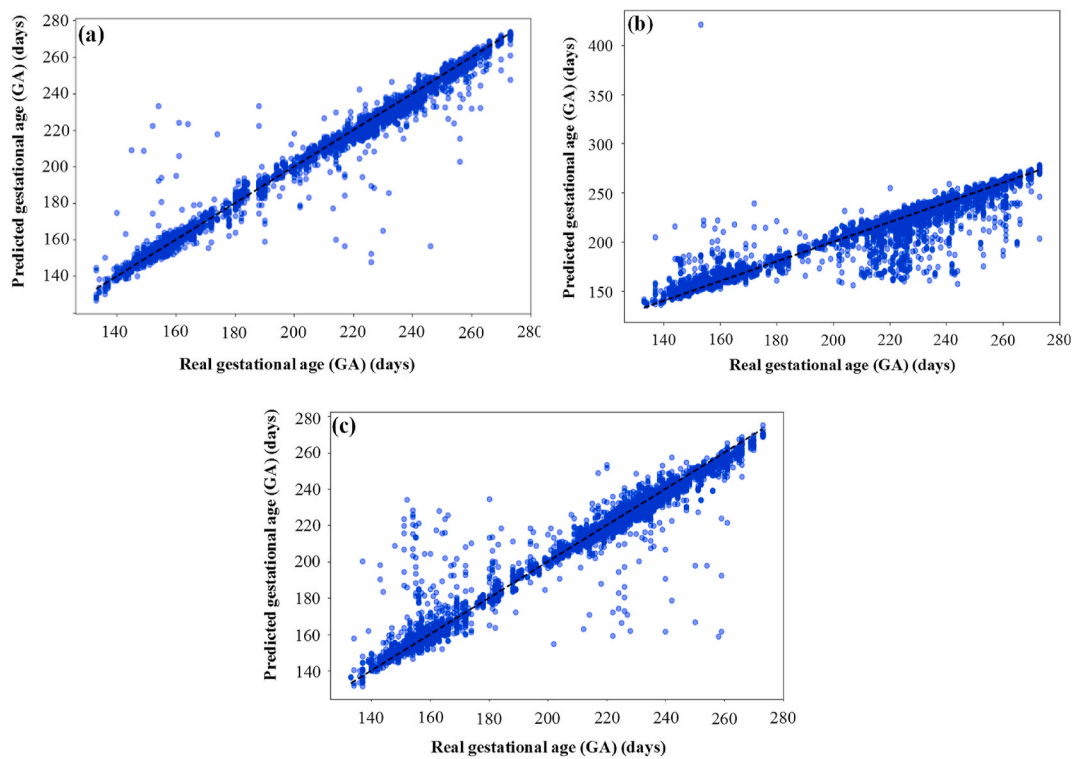
Fig. 9. Actual and predicted GA plots of a single plane for the proposed model Xception-MHA (8-32-32)-Dense (512)-Dense (1) (a) axial, (b) coronal and (c) sagittal images for predicting gestational age in the fetal brain MRI dataset.

the actual GA values, confirming the model's ability to predict the GA from sagittal MR images with a reasonable degree of accuracy. The sagittal plane performed moderately well, with an R-squared value of 93.6, an MAE of 4.49 days, and a PCC of 96.7. The sagittal plane images likely provide a good balance of features that are distinct enough for accurate GA prediction but may still have some variability compared with the axial plane. Overall, Fig. 9 demonstrates the effectiveness of the proposed model in predicting GA from fetal brain MR images, with the axial plane providing the highest accuracy, followed by the sagittal and then the coronal planes.

4.3.4. Time complexity analysis

Time complexity analysis is a crucial aspect of evaluating DL models, as it provides insight into the computational resources required for training and inference.

Table 9 presents a comparison of complexity across several models with the proposed model, focusing on parameters (million), FLOPs (Floating Point Operations - Giga), and runtime (seconds). The proposed model has 71.92 million parameters and requires 1.19 GFLOPs, which indicates moderate complexity compared with the others reported in the literature. However, its runtime of 1.95 s is substantially longer than those of the other models, indicating that while it balances the parameter count and computational load efficiently, it may be less optimized for speed. This suggests a trade-off in the proposed model between complexity and execution time, potentially due to more intricate processing steps or deeper network architecture.

Table 9
Comparison of algorithm complexity.

Reference	Parameters (M)	FLOPs (G)	Runtimes (S)
Shen et al. [16]	32.07	0.95	0.13
Liao et al. [35]	117.81	2.75	0.20
Feng et al. [22]	48.95	1.34	0.14
Proposed model	71.92	1.19	1.95

4.4. Performance comparison with SOTA models

Table 10 illustrates the superior performance of the proposed Xception-MHA-Dense model in predicting GA from fetal brain MR images compared with existing SOTA models. Kojita et al. [19] utilized the VGG-16 model on a smaller dataset of 184 patients, yielding an MAE of 5.8 weeks (approximately 40.6 days), but did not report R^2 or PCC values. Additionally, Kojita et al. used only a pretrained model without custom modifications, and the limited number of images further contributed to their higher error rates. Diego et al. [12] employed a 50-KNF-RF model with 243 patients, achieving an MAE of 2.22 weeks (approximately 15.54 days) but also lacked R^2 and PCC metrics. Compared with modern DL techniques, this traditional machine learning approach resulted in poor performance. Wen et al. [25] and Liao et al. [35] used attention-based and multibranch deformable CNN models, respectively, reporting high R^2 values (92 % and 94.7 %) and moderate MAE values expressed in weeks. Although the results of Wen et al. were impressive, with an R^2 of 92 % and an MAE of 0.767 weeks (approximately 5.37 days), the proposed model still outperformed them. Shen et al. [16,20] applied attention-guided ResNet-50 models on larger datasets, achieving R^2 values of approximately 94–94.5 %, MAEs of 6.7–6.8 days and PCCs of up to 97 %. In the latest paper by Shen, they used the same dataset as this study but achieved an R^2 of 94.5 %, whereas the proposed model reached 96.5 %, with a lower MAE of 3.8 days, indicating superior accuracy and robustness. By leveraging a robust architecture combining Xception with multihead attention and dense layers, which was applied to an extensive dataset of 741 patients and 52,900 images, the proposed model demonstrated the highest R^2 value of 96.5 %, the lowest MAE of 3.8 days, and an impressive PCC of 98.5 %, confirming its superior accuracy and reliability in GA prediction. For segmentation tasks, studies such as Tran et al. [36] and Tuhinangshu et al. [37] achieved high accuracy. Tran et al. reported median Dice scores of 0.95 for intracranial contents and 0.90 for brain segmentation. Tuhinangshu et al. achieved a Jaccard similarity of 77 % and a Dice score of 82 %. While these segmentation results are impressive,

Table 10
Comparative analysis between the proposed Xception-MHA-Dense model and other state-of-the-art models.

Reference	Models	Num. of patients	Num. of images	R^2	MAE	PCC
Kojita et al. [19]	VGG-16	184	184	–	5.8 (weeks)	–
Diego et al. [12]	50-KNF-RF	243	–	51	2.22 (weeks)	–
Wen et al. [25]	Attention based Residual Network	–	659	92	0.767 (weeks)	–
Liao et al. [35]	Multibranch deformable CNN	289	–	94.7	0.751 (weeks)	–
Shen et al. [20]	Attention guided ResNet-50	741	1927	94	6.8 (days)	–
Shen et al. [16]	Attention guided ResNet-50	741	52,900	94.5	6.7 (days)	97
Zhao et al. [18]	Attention-based hemispheric relation inference network	531	–	89	0.53 (weeks)	94
Ran et al. [21]	JoCoRank algorithm	157	1327	93	0.693 (weeks)	–
Ziteng et al. [22]	PDFF-CNN	157	1327	90.4	0.84 (weeks)	–
Proposed model	Xception-MHA (8-32-32)-Dense (512)-Dense (1)	741	52,900	96.5	3.8 (days)	98.5

they focus primarily on ROI segmentation rather than direct GA prediction.

Zhao et al. [18] used an attention-based hemispheric relation inference network, and Ziteng et al. [22] PDFF-CNN, which predicts GAs in the narrower range of 22–24 weeks, achieved R^2 values of 89 % and 90.4 %, with MAEs of 0.53 weeks and 0.84 weeks, respectively. In contrast, our model predicts the GA across a broader range of 19–38 weeks, demonstrating greater stability and robustness. Additionally, Ran et al. [22] applied the JoCoRank algorithm to a dataset of 157 patients and achieved an R^2 of 93 % and an MAE of 0.693 weeks, which are smaller compared with our model. The proposed Xception-MHA-Dense model, with data from 741 patients and 52,900 images, achieves an impressive R^2 of 96.5 %, an MAE of 3.8 days, and a PCC of 98.5 %. This broader age prediction range and higher accuracy highlight the model's ability to handle more variation effectively, establishing it as a more reliable and stable solution for GA prediction. Compared with existing models, the proposed Xception-MHA-Dense model excels in GA prediction from fetal brain MR images, offering a higher R^2 value, lower MAE, and higher PCC. The proposed method combines advanced DL techniques, multiplanar imaging, and a larger dataset, addressing the limitations of previous studies and setting a new benchmark for accuracy and reliability in GA prediction.

4.5. Discussion

In the discussion section, the experimental results have been interpreted, considering both the strengths and limitations of the proposed model. The potential applications of the proposed model findings and directions for future research have been explored.

4.5.1. Strengths, limitations and future work

Predicting GA from fetal brain MR images is a challenging task, with limited research in this area due to its complexity. However, this study successfully addresses this challenge by employing a novel DL model that outperforms SOTA models. The proposed model leverages multiplanar images, significantly enhancing its predictive accuracy because MHA might allow the capture of diverse anatomical perspectives present in MR images, contributing to more robust predictions of the GA. To enhance the discriminative power of the extracted features, a dense layer was incorporated into the model architecture. This layer might have enabled nonlinear transformations, facilitating better regression performance and improving the overall predictive accuracy of the model. A key strength of this study is its robustness, demonstrated by the extensive dataset used, comprising 52,900 images from 741 patients. This large dataset ensures the model's reliability and generalizability. Furthermore, the proposed model predicts the GA in days rather than in weeks, providing more precise and specific estimations. This level of granularity is crucial because it handles a wider range of variations, from 140 to 280 days, than the narrower range of 19–39 weeks used in other studies. This ability to predict GA in days highlights the model's capacity to manage more intricate variations in fetal development. This is particularly important for clinical applications where precise timing can influence critical decisions in prenatal care. The proposed model outperformed existing SOTA models by producing the lowest MAE of 3.8 days. Overall, the proposed model's high accuracy, extensive dataset, and fine-grained predictions underscore its robustness and potential for significant clinical impact.

Despite its strengths, the proposed model has several limitations. The prediction time for an image is relatively high for the proposed model. Notably, the image dataset used to develop the model was taken by instruments operating at magnetic field strengths of both 1.5 T and 3 T, but individual images are not specifically labeled for the Tesla values. The variation in magnetic field strength during image creation can indeed impact image quality, potentially affecting model performance. Additionally, the unavailability of datasets with clearly documented Tesla values restricts our ability to comprehensively evaluate the model's robustness across diverse imaging conditions. While it outperformed all other SOTA models when the GA was predicted via multiplane images, its performance varied significantly across different planes. The model showed impressive results when predicting the GA from axial plane images and performed well with sagittal plane images. However, its performance was notably poorer with coronal plane images, indicating that the model struggles to accurately learn the variations in coronal images. This limitation affects the results of two-plane combinations involving coronal images, such as coronal-sagittal and coronal-axial images, which do not achieve as much accuracy as other combinations do. The inability to predict the GA accurately from coronal images suggests a critical area for future improvement. Enhancing the model's ability to handle coronal images more effectively will be essential for achieving consistently high performance across all imaging planes and improving the overall robustness of GA predictions.

4.5.2. Clinical significance

The clinical significance of this research lies in its potential to greatly enhance prenatal care through the accurate prediction of GA via the application of DL models to fetal brain MR images. DL models provide a reliable and highly accurate method for predicting GAs, surpassing traditional manual methods. This improved accuracy is crucial for making informed obstetric decisions, such as determining the optimal timing for interventions such as magnesium sulfate administration for neuroprotection or steroids for fetal lung maturation. Traditional methods for estimating GA from MRI data involve manual processes that require radiologists to measure various brain structures and compare these measurements to established growth charts [38]. This manual process is time-consuming, prone to inter-observer variability, and often less accurate, as it relies heavily on the experience and expertise of the clinician [38,39]. With precise GA predictions, clinicians can better identify and manage high-risk pregnancies, ensuring timely interventions that can significantly improve maternal and fetal outcomes.

Accurate GA estimation allows for early identification of potential developmental issues, facilitating prompt and appropriate medical responses. At the neonatal level, precise GA estimation helps in identifying vulnerable neonates, such as those who are small for GA or who are preterm, ensuring that they receive the necessary care. Accurate GA predictions from fetal MRI can enhance the management and outcomes of these infants by providing better insights into their developmental status. Preterm birth is a leading cause of mortality in children under five years of age [39]. Accurate GA estimation using DL models can help in understanding the prevalence of preterm births and implementing effective strategies to mitigate associated risks. This can contribute to better monitoring and intervention strategies during pregnancy, ultimately reducing complications related to preterm birth. The use of DL models for GA prediction can automate and streamline clinical workflows, reducing the reliance on manual assessments. This automation can save time and resources, allowing for more efficient prenatal screening and monitoring practices. Clinicians can focus more on patient care than on time-consuming manual calculations, thereby increasing overall healthcare efficiency. DL models offer scalability and consistency in GA predictions, ensuring uniformity in assessments across different healthcare settings. This consistency reduces the variability and potential errors associated with manual methods, leading to more reliable and standardized prenatal care. Overall, the integration of DL in GA prediction from fetal brain MR images represents a significant advancement in prenatal care, offering precise, efficient, and scalable solutions that improve clinical decision-making and patient outcomes.

5. Conclusions

This research demonstrated the efficacy of a novel DL model in predicting GA from fetal brain MR images, providing a robust and accurate method that surpasses existing state-of-the-art models. The proposed model, which leverages a combination of Xception with MHA and dense layers, exhibits high accuracy and robustness across various imaging planes. The proposed model achieved promising results with 8 attention heads, a 32-dimensional key space, and a 32-dimensional value space, attaining an R^2 value of 96.5 %, a mean absolute error of 3.80 days, and a PCC of 98.50 % on the test set. Additionally, the 5-fold cross-validation results reinforce the model's reliability, with an average R^2 of 95.94 %, an MAE of 3.61 days, and a PCC of 98.02 %, demonstrating consistently high performance across different subsets of data. This high performance can be attributed to the complementary information provided by multiplane combinations, enhancing the model's ability to capture relevant features for accurate GA prediction. The analysis revealed that the multiplanar approach outperformed the combination of any two planes, such as the axial-sagittal combination (R^2 95.7 %, MAE 4.97 days, PCC 98.4 %), the axial-coronal combination (R^2 90.6 %, MAE 5.11 days, PCC 95.4 %) and the sagittal-coronal combination (R^2 88.2 %, MAE 7.52 days, PCC 94.0 %). Among single planes, the axial plane provided the best performance (R^2 97.1 %, MAE 3.24 days, PCC 98.6 %), followed by the sagittal (R^2 93.6 %, MAE 4.49 days, PCC: 96.7 %) and coronal planes (R^2 86.3 %, MAE 7.06 days, PCC 93.0 %). This model outperformed all other state-of-the-art models. The proposed DL method can automate the GA estimation process with high accuracy, enabling precise, efficient, and scalable solutions that improve clinical decision-making and patient outcomes and marking a significant step forward in prenatal care. In the future, further improvements in the DL model with higher computational accuracy and efficiency for real-time clinical applications will be implemented.

CRediT authorship contribution statement

Mohammad Asif Hasan: Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Fariha Haque:** Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Tonmoy Roy:** Writing –

original draft, Methodology, Investigation, Data curation. **Mahedi Islam:** Writing – original draft, Methodology, Investigation, Data curation. **Md Nahiduzzaman:** Writing – review & editing, Validation, Supervision, Formal analysis. **Mohammad Mahedi Hasan:** Writing – original draft, Methodology, Data curation. **Mominul Ahsan:** Writing – review & editing, Visualization, Validation, Supervision, Formal analysis. **Julfikar Haider:** Writing – review & editing, Visualization, Validation, Supervision, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A.F. Alexander-Bloch, A. Raznahan, S.N. Vandekar, J. Seidlitz, Z. Lu, S.R. Mathias, E. Knowles, J. Mollon, A. Rodrigue, J.E. Curran, Imaging local genetic influences on cortical folding, Proc. Natl. Acad. Sci. USA 117 (13) (2020) 7430–7436.

[2] T. Rogne, A.A. Engstrøm, G.W. Jacobsen, J. Skranes, H.F. Østgård, M. Martinussen, Fetal growth, cognitive function, and brain volumes in childhood and adolescence, Obstet. Gynecol. 125 (3) (2015) 673–682.

[3] H.J. Yun, L. Vasung, T. Tarui, C.K. Rollins, C.M. Ortinau, P.E. Grant, K. Im, Temporal patterns of emergence and spatial distribution of sulcal pits during fetal life, Cerebr. Cortex 30 (7) (2020) 4257–4268.

[4] C.R. Madan, E.A. Kensinger, Predicting age from cortical structure across the lifespan, Eur. J. Neurosci. 47 (5) (2018) 399–416.

[5] J.H. Cole, J. Raffel, T. Friede, A. Eshaghi, W.J. Brownlee, D. Chard, N. De Stefano, C. Enzinger, L. Pirpamer, M. Filippi, Longitudinal assessment of multiple sclerosis with the brain-age paradigm, Ann. Neurol. 88 (1) (2020) 93–105.

[6] N.N. Andescavage, A. Du Plessis, R. McCarter, A. Serag, I. Evangelou, G. Vezina, R. Robertson, C. Limperopoulos, Complex trajectories of brain development in the healthy human fetus, Cerebr. Cortex 27 (11) (2017) 5274–5283.

[7] H.J. Yun, J.D.R. Perez, P. Sosa, J.A. Valdés, N. Madan, R. Kitano, S. Akiyama, B. G. Skotko, H.A. Feldman, D.W. Bianchi, Regional alterations in cortical sulcal depth in living fetuses with down syndrome, Cerebr. Cortex 31 (2) (2021) 757–767.

[8] A. Conde-Agudelo, R. Romero, Antenatal magnesium sulfate for the prevention of cerebral palsy in preterm infants less than 34 weeks' gestation: a systematic review and metaanalysis, Am. J. Obstet. Gynecol. 200 (6) (2009) 595–609.

[9] L. Liu, S. Oza, D. Hogan, Y. Chu, J. Perin, J. Zhu, J.E. Lawn, S. Cousens, C. Mathers, R.E. Black, Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals, Lancet 388 (10063) (2016) 3027–3035.

[10] A.C. Lee, P. Panchal, L. Folger, H. Whelan, R. Whelan, B. Rosner, H. Blencowe, J. E. Lawn, Diagnostic accuracy of neonatal assessment for gestational age determination: a systematic review, Pediatrics 140 (6) (2017).

[11] Y. Lu, X. Fu, F. Chen, K.K. Wong, Prediction of fetal weight at varying gestational age in the absence of ultrasound examination using ensemble learning, Artif. Intell. Med. 102 (2020) 101748.

[12] D. Fajardo-Rojas, M. Hall, D. Cromb, M.A. Rutherford, L. Story, E. Robinson, J. Hutter, Predicting gestational age at birth in the context of preterm birth from multi-modal fetal MRI, medRxiv (2024), 2024.02. 17.24302791.

[13] C. Ross, P. Deruelle, M. Pontvianne, L. Lecointre, S. Wieder, P. Kuhn, M. Lodi, Prediction of adverse neonatal adaptation in fetuses with severe fetal growth restriction after 34 weeks of gestation, Eur. J. Obstet. Gynecol. Reprod. Biol. 296 (2024) 258–264.

[14] M. Alzubaidi, M. Agus, U. Shah, M. Makhlouf, K. Alyafei, M. Househ, Ensemble transfer learning for fetal head analysis: from segmentation to gestational age and weight prediction, Diagnostics 12 (9) (2022) 2229.

[15] L.H. Lee, E. Bradburn, R. Craik, M. Yaqub, S.A. Norris, L.C. Ismail, E.O. Ohuma, F. C. Barros, A. Lambert, M. Carvalho, Machine learning for accurate estimation of fetal gestational age based on ultrasound images, NPJ Digital Medicine 6 (1) (2023) 36.

[16] L. Shen, J. Zheng, E.H. Lee, K. Shpanskaya, E.S. McKenna, M.G. Atluri, D. Plasto, C. Mitchell, L.M. Lai, C.V. Guimaraes, Attention-guided deep learning for gestational age prediction using fetal brain MRI, Sci. Rep. 12 (1) (2022) 1408.

[17] L. Shen, J. Zheng, E.H. Lee, K. Shpanskaya, E.S. McKenna, M.G. Atluri, D. Plasto, C. Mitchell, L.M. Lai, C.V. Guimaraes, H. Dahmoush, J. Chueh, S.S. Halabi, J. M. Pauly, L. Xing, Q. Lu, O. Oztekin, B.M. Kline-Fath, K.W. Yeom, Attention-guided deep learning for gestational age prediction using fetal brain MRI, Sci. Rep. 12 (1) (Jan 26 2022) 1408.

[18] L. Zhao, D. Zhu, X. Wang, X. Liu, T. Li, B. Wang, Z. Yao, W. Zheng, B. Hu, An attention-based hemispheric relation inference network for perinatal brain age prediction, IEEE Journal of Biomedical Health Informatics (2024).

[19] Y. Kojita, H. Matsuo, T. Kanda, M. Nishio, K. Sofue, M. Nogami, A.K. Kono, M. Hori, T. Murakami, Deep learning model for predicting gestational age after the first trimester using fetal MRI, Eur. Radiol. 31 (6) (Jun 2021) 3775–3782.

[20] L. Shen, K. Shpanskaya, E. Lee, E. McKenna, M. Maleki, Q. Lu, S. Halabi, J. Pauly, K. Yeom, Deep learning with attention to predict gestational age of the fetal brain, arXiv preprint arXiv:1812.07102 (2018).

[21] R. Zhou, Y. Liu, W. Xia, Y. Guo, Z. Huang, H. Gan, A. Fenster, JoCoRank: joint correlation learning with ranking similarity regularization for imbalanced fetal brain age regression, Comput. Biol. Med. 171 (2024) 108111.

[22] Z. Feng, R. Zhou, W. Xia, S. Wang, Y. Liu, Z. Huang, H. Gan, PDFF-CNN: an attention-guided dynamic multi-orientation feature fusion method for gestational age prediction on imbalanced fetal brain MRI dataset, Med. Phys. 51 (5) (2024) 3480–3494.

[23] M. Liu, M. Lu, S.Y. Kim, H.J. Lee, B.A. Duffy, S. Yuan, Y. Chai, J.H. Cole, X. Wu, A. W. Toga, Brain age predicted using graph convolutional neural network explains neurodevelopmental trajectory in preterm neonates, Eur. Radiol. 34 (6) (2024) 3601–3611.

[24] A. Ansari, K. Pillay, L. Baxter, E. Arasteh, A. Dereymaeker, G.S. Mellado, K. Jansen, G. Naulaers, A. Bhatt, S. Van Huffel, Brain age as an estimator of neurodevelopmental outcome: a deep learning approach for neonatal cot-side monitoring, bioRxiv (2023) 2023, 01. 24.525361.

[25] W. Shi, G. Yan, Y. Li, H. Li, T. Liu, C. Sun, G. Wang, Y. Zhang, Y. Zou, D. Wu, Fetal brain age estimation and anomaly detection using attention-based deep ensembles with uncertainty, Neuroimage 223 (2020) 117316.

[26] J.-H. Kim, J. De Asis-Cruz, K.M. Cook, C. Limperopoulos, Gestational age-related changes in the fetal functional connectome: in utero evidence for the global signal, Cerebr. Cortex 33 (5) (2023) 2302–2314.

[27] M. Mazher, A. Qayyum, D. Puig, M. Abdel-Nasser, Effective approaches to fetal brain segmentation in MRI and gestational age estimation by utilizing a multiview deep inception residual network and radiomics, Entropy 24 (12) (2022) 1708.

[28] C. Calixto, F. Machado-Rivas, D. Karimi, M.C. Cortes-Albornoz, L.M. Acosta-Buitrago, S. Gallo-Bernal, O. Afacan, S.K. Warfield, A. Gholipour, C. Jaimes, Detailed anatomic segmentations of a fetal brain diffusion tensor imaging atlas between 23 and 30 weeks of gestation, Hum. Brain Mapp. 44 (4) (2023) 1593–1602.

[29] F. Vahedifard, X. Liu, K.K. Marathu, M. Kocak, H.A. Ai, M.P. Supanich, S. Adler, S. M. Ansari, M. Akyuz, J.O. Adepoju, Artificial intelligence prediction of gestational age of fetal in brain magnetic resonance imaging versus ultrasound using three different biometric measurements, Preprints (2023).

[30] G. Hong, X. Chen, J. Chen, M. Zhang, Y. Ren, X. Zhang, A multi-scale gated multi-head attention depthwise separable CNN model for recognizing COVID-19, Sci. Rep. 11 (1) (2021) 18048.

[31] H.T. Pham, J. Awange, M. Kuhn, Evaluation of three feature dimension reduction techniques for machine learning-based crop yield prediction models, Sensors 22 (17) (Sep 1 2022).

[32] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.

[33] M.R. Islam, M. Nahiduzzaman, M.O.F. Goni, A. Sayeed, M.S. Anower, M. Ahsan, J. Haider, Explainable transformer-based deep learning model for the detection of malaria parasites from blood cell images, Sensors 22 (12) (Jun 8 2022).

[34] M. Lipa, R.K. Pooh, M. Wielgoś, Three-dimensional neurosonography—a novel field in fetal medicine, Ginekol. Pol. 88 (4) (2017) 215–221.

[35] L. Liao, X. Zhang, F. Zhao, J. Lou, L. Wang, X. Xu, H. Zhang, G. Li, Multi-branch deformable convolutional neural network with label distribution learning for fetal brain age prediction, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 424–427.

[36] C. Tran, P. Nedelec, D. Weiss, J. Rudie, L. Kini, L. Sugrue, O. Glenn, C. Hess, A. Rauschecker, Development of gestational age–based fetal brain and intracranial volume reference norms using deep learning, Am. J. Neuroradiol. 44 (1) (2023) 82–90.

[37] T. Gangopadhyay, S. Halder, P. Dasgupta, K. Chatterjee, D. Ganguly, S. Sarkar, S. Roy, MTSE U-Net: an architecture for segmentation, and prediction of fetal brain and gestational age from MRI of brain, Network Modeling Analysis in Health Informatics Bioinformatics 11 (1) (2022) 50.

[38] U. Vovk, F. Pernus, B. Likar, A review of methods for correction of intensity inhomogeneity in MRI, IEEE Trans. Med. Imag. 26 (3) (2007) 405–421.

[39] J. Hong, H.J. Yun, G. Park, S. Kim, Y. Ou, L. Vasung, C.K. Rollins, C.M. Ortinau, E. Takeoka, S. Akiyama, Optimal method for fetal brain age prediction using multiplanar slices from structural magnetic resonance imaging, Front. Neurosci. 15 (2021) 714252.