ORIGINAL ARTICLE

MTSE U-Net: an architecture for segmentation, and prediction of fetal brain and gestational age from MRI of brain

Tuhinangshu Gangopadhyay[1] · Shinjini Halder[1] · Paramik Dasgupta[2] · Kingshuk Chatterjee[3] · Debayan Ganguly[1] · Surjadeep Sarkar[1] · Sudipta Roy[4]

Abstract
Fetal brain segmentation and gestational age prediction have been under active research in the field of medical image processing for a long time. However, both these tasks are challenging due to factors like difficulty in acquiring a proper fetal brain image owing to the fetal movement during the scan. With the recent advancements in deep learning, many models have been proposed for performing both the tasks, individually, with good accuracy. In this paper, we present Multi-Tasking Single Encoder U-Net, MTSE U-Net, a deep learning architecture for performing three tasks on fetal brain images. The first task is the segmentation of the fetal brain into its seven components: intracranial space and extra-axial cerebrospinal fluid spaces, gray matter, white matter, ventricles, cerebellum, deep gray matter, and brainstem, and spinal cord. The second task is the prediction of the type of the fetal brain (pathological or neurotypical). The third task is the prediction of the gestational age of the fetus from its brain. All of this will be performed by a single model. The fetal brain images can be obtained by segmenting it from the fetal magnetic resonance images using any of the previous works on fetal brain segmentation, thus showing our work as an extension of the already existing segmentation works. The Jaccard similarity and Dice score for the segmentation task by this model are 77 and 82%, respectively, accuracy for the type of prediction task is 89% and the mean absolute error for the gestational age task is 0.83 weeks. The salient region identification by the model is also tested and these results show that a single model can perform multiple, but related, tasks simultaneously with good accuracy, thus eliminating the need to use separate models for each task.

Keywords Medical image processing · Fetal brain segmentation · Fetal gestational age prediction · Deep learning · Convolutional neural networks

✉ Sudipta Roy
  sudipta1.roy@jioinstitute.edu.in

  Tuhinangshu Gangopadhyay
  tuhinangshu.ganguly@gmail.com

  Shinjini Halder
  shinjini.halder01@gmail.com

  Paramik Dasgupta
  paramikdasgupta@gmail.com

  Kingshuk Chatterjee
  kingshukchatterjee@gcect.ac.in

  Debayan Ganguly
  debayan@gcelt.org

  Surjadeep Sarkar
  surjadeep@gcelt.org

[1] Government College of Engineering and Leather Technology, Kolkata 700106, India

[2] Asian Institute of Technology, Khlong Nueng, Thailand

[3] Government College of Engineering and Ceramic Technology, Kolkata 700010, India

[4] Artificial Intelligence & Data Science, Jio Institute, Navi Mumbai 410206, India

1 Introduction

Computer science plays an essential role in many fields of study including medical science. The advancement of artificial intelligence and machine learning can assist doctors in identifying anomalies in the future. Ultrasonography (USG) is the primary method for fetal imaging and diagnosis of fetal abnormalities, but fetal magnetic resonance imaging (MRI) has been proven to be advantageous over USG because of better image contrast of soft tissues, larger view, and 3D visualization (Levine 2001) and is also not affected by maternal obesity and severe oligohydramnios (Roy et al. 2017a). MRI is non-invasive medical imaging that uses strong magnetic fields and radio waves rather than ionization radiation to produce detailed images of the organs. Iman A Hosny and Hamed S Elghawabi (Hosny and Elghawabi 2010) concluded that ultrasound being the primary choice for screening method, MRI can be used as a "complementary adjunctive modality" to ultrasonography and has been proved to be effective in determining many abnormalities in a growing fetus. Human brain development begins in the third gestational week and thereafter, under genetic influence, it shows rapid anatomical changes like the formation of characteristic gyri and sulci, an increase in brain volume, and neurological developments like neurons producing chemical signals for communication (Konkel 2018; Chi et al. 1977; Roy and Bandyopadhyay 2016). It is believed that the communication established among the neurons in the brain affects personality, behavior, cognition, and memory of an individual (Gale 2004) and any deviation from standard development patterns is a pathological condition like schizophrenia, autism, attention deficit hyperactivity disorder, congenital heart diseases (Murray et al. 1991; Scher 2003 Jun; Levman and Takahashi 2015; Jaimes et al. 2020), etc.

Machine learning recognizes and classifies patterns and reveals intricate details in data. After its introduction in medical imaging, different machine learning models are trained to make predictions of diagnosis. Advancements in hardware configurations and machine learning led to the creation of deep learning. Deep learning, which is a subset of machine learning, tries to mimic the human brain. It is composed of many layers and multiple hidden layers; these layers are formed of nodes that work collectively to produce a result; this whole process resembles a neural network. Deep neural networks are capable of feature extraction, that is reducing the number of input features and selecting only those features that accurately and concisely explain our desired results from unstructured data, without human interventions, subsequently by backtracking and gradient descent, thus increasing the accuracy of predictions. Large training data, domain-specific knowledge, and developing algorithmic techniques for different scanning protocols can improve the result obtained from deep learning (Shen et al. 2017).

The downside of deep learning is its need for large datasets to produce high-accuracy results. Its dependency on large datasets is proportional to the complexity of the images and image set. On the other hand, conventional image processing relies on manual extraction of features, albeit this method is labor intensive but is often more effective when it comes to small datasets. Lufan Liao et al. (Liao, et al. 2020) addressed the problem using label distribution learning (LDL). Hagerty et al. (Hagerty et al. 2017) discussed that a fusion of deep learning with manual feature extraction can reduce the need for large datasets and produce better results. Rajchl et al. (Rajchl et al. 2016) employed a crowdsourcing platform for weak annotation by non-experts and then applied it to a fully convolutional neural network and pointed out that the result obtained was not as accurate as annotations by experts but fair enough. N. Khalili et al. (Khalili et al. 2019) proposed a segmentation method that uses CNN twice, once to extract intracranial volume, and next to segment the extracted volume into seven types of brain tissues.

Identifying early signs and providing early treatment can improve or cure a person's health. Image segmentation can detect suspicious regions from medical images, and this may foretell many diseases like brain tumors and breast cancer (Joseph 2014; Roy et al. 2022; Roy and Shoghi 2019). Ultrasound is the major medical in-utero imaging technique used to examine a developing fetus, but it generates 2D images which in contrast to 3D MRI, does not allow visualization from all perspectives. So, any suspected diagnosis in USG is further examined by MRI. Fetal MRI faces some challenges, like the arbitrary orientation of the fetus, the small size of the fetus, and fetal motion, substantially during early gestation as these add noise to data (Roy et al. 2017b). Brain extraction and super-resolution reconstruction play an important role in the study of MRI images. A significant image segmentation process called U-Net was developed by Ronneberger et al. (Ronneberger et al. 2015). It was built upon a contracting path to capture context and a symmetric expanding path that enables precise localization. Further modifications of U-Net include 3D U-Net (Çiçek et al. 2016) which extends the 2D U-Net segmentation into its equivalent 3D segmentation, U-Net++ (Zhou et al. 2018), generated by appending a dense convolution block and having dense skip connections on skip pathways, and also deep supervision for improvement of the model, etc. A variation by Rampun et al. (Rampun et al. 2019) in contrast to the original U-Net includes more convolutional blocks, a combination of balanced cross-entropy, exponential linear unit, and 'RMSprop' for learning generalization, and a drop-out layer at the end of the last three blocks in the downsampling

part. U-Net has extensive use in the field of medical science and has been used for fetal brain extraction from MRI slices (Lou et al. 2019; Salehi et al. 2018), tumor segmentation (Do et al. 2021; Failed 2019).

Gestational age can also be derived from the fetal MRI scans. To calculate accurate gestational age from MRI images in the second half of gestation, Yuequan Shi et al. (Shi et al. 2020) established a second-order polynomial regression model to be the best descriptor of biometric measure while a linear model accurately predicted gestational age in the second and third trimesters. A work by Liyue Shen et al. (2022) proposed an attention-guided, multi-view deep learning network that analyzes MRI accurately to predict gestational age. The proposed regression algorithm provides an automated machine-enabled tool with the potential to better characterize in-utero neurodevelopment and guide real-time gestational age estimation after the first trimester.

In this paper, we developed Multi-Tasking Single Encoder U-Net (abbreviated as MTSE U-Net), a deep learning-based architecture, to solve three tasks. The first task is to segment the fetal brain into seven major components: intracranial space and extra-axial cerebrospinal fluid spaces (CSF), gray matter (GM), white matter (WM), ventricles (LV), cerebellum (CBM), deep gray matter, especially thalamus and putamen (SGM), and brainstem and spinal cord (BS). The second task is to predict the type of the fetal brain ('Pathological' or 'Neurotypical') and the third task is to predict the gestational age of the fetus from its brain image. Thus, this model can be interpreted as a 'many-in-one' model, which can perform multiple, but related, tasks (suggesting a model for three tasks is referred to as 'the combined task' throughout the rest of the paper). We have compared the performance of our model with some of the previously proposed works on these tasks. This model is a modified version of the U-Net architecture (Ronneberger et al. 2015). The motivation behind our work is:

1. U-Net is used as the base model because it can encode the necessary spatial and orientation data, which can later be decoded to return a precise segmentation. It can precisely localize borders and is found to be efficient in several medical imaging segmentation use cases.
2. Developing a model for the combined task is inspired by the fact that, since all the tasks require the same input (thus being somewhat related) and the model is required to learn to encode the essential features at the training step for all of them; thus, the encoder can be kept common, trying to use the same encoded information for all the three tasks.
3. Our work can be seen as an extension of the previous papers on segmentation of the fetal MRI to return the fetal brain image, after eliminating the neighboring

components like the skull surrounding the brain and the womb of the mother.

Our work made the following contributions:

1. We have proposed the MTSE U-Net model for performing the combined task with a high accuracy.
2. As per our knowledge, this is the first time that someone has tried to perform multiple tasks with a single model in the field of medical image processing with fetal brain MRI.
3. We have proceeded a step further in making use of the segmented fetal brain as input and segmenting it into its seven major components, predicting the type of the brain and finally the gestational age of the fetus from its brain, all at the same time, by the same model.
4. We have compared the performance of our model with three state-of-the-art segmentation models: Deep U-Net (Rampun et al. 2019), FCN (Rajchl et al. 2016) and U-Net (Ronneberger et al. 2015).

2 Materials and methods

At first, we applied some pre-processing steps to the input for convenience and easier interpretation of data. It is followed by developing and experimenting with several modified versions of the U-Net model for the combined task, comparing their outputs, and selecting the best model out of them, as discussed further in detail. We have used Google Colaboratory for this work.

2.1 Data pre-processing

The data pre-processing and data cleaning steps necessary for our model are:

1. To increase training time and space efficiency and for easy accessibility, the 3D dataset (details about the dataset are available in the 'data availability' section) is divided into a set of 2D images along all three axes. An example, the input image, and its corresponding segmentation are shown in Fig. 1a and b. The 3D dataset produces a total of (256*3*80=) 61,440 images, but this immense number of images for training will demand a huge training time and might even lead to the overfitting of the model. Thus, we have used around 14,000 images (13,500 for training and validation, and 550 for testing) from all the MRIs. We chose images from all the fetus samples to be able to capture every variation in the fetal brain for an accurate prediction of all the three tasks. This is because the dataset has fetal MRI of both neurotypical and pathological types (which vary greatly

Fig. 1 An example image along with its segmentation and all the pre-processing steps applied on it; except for **c**, which shows a different (and already blurred) image
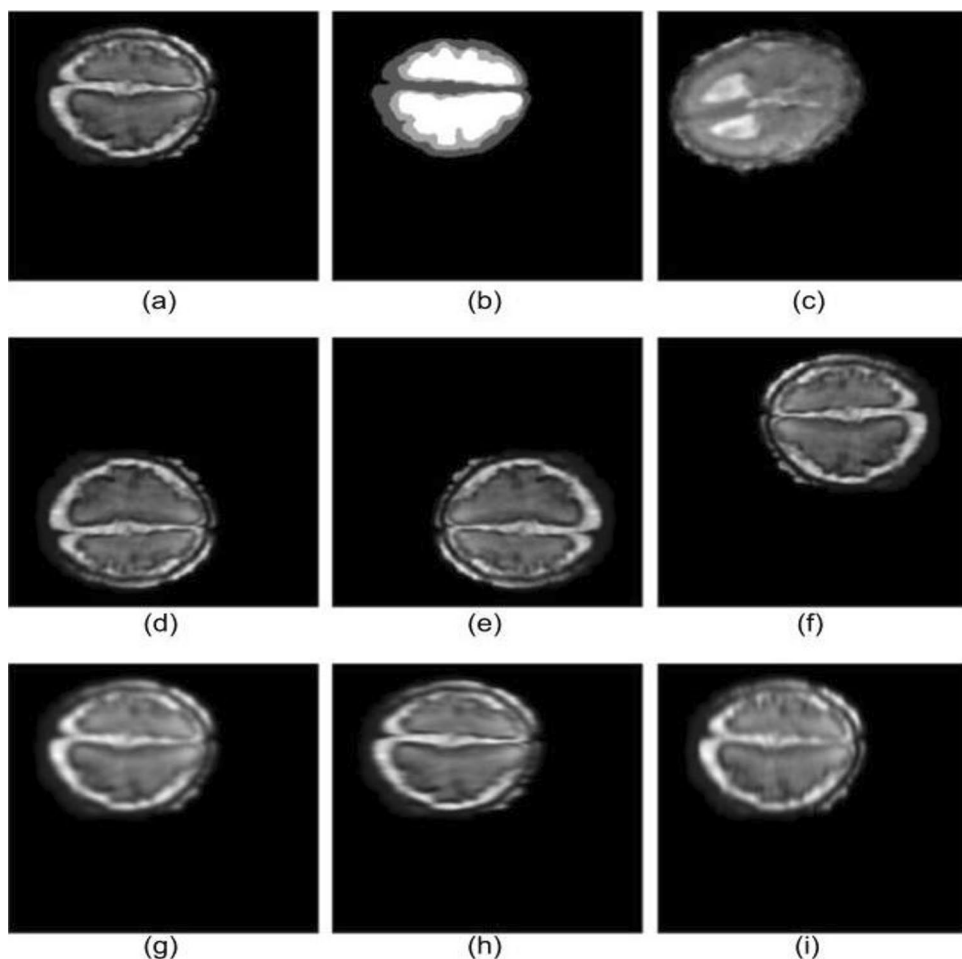
among themselves in their structures depending on the condition the brain is suffering from), with the fetus gestational age ranging from 20 to 35 weeks (which is the duration of fast paced brain development). As a result, our model is capable of working with any type of fetal brain.

2. These images are not chosen arbitrarily, rather, after removing the images/slices having no brain part in them (i.e., completely filled with zeroes), we have chosen a number of equally separated slices for each 3D sample and along each axis. The total number of slices (n) chosen from all the samples combined can be represented by the following equation:

$$n = \sum_{j=1}^{80} \sum_{i=1}^{3} \left\lfloor \frac{(m - z_{ij})}{\left\lfloor \frac{(m - z_{ij})}{k} \right\rfloor} \right\rfloor$$

where, first summation is the aggregation all the samples, the second summation is the sum over all the three axes, $m = 256$ (since all the samples, along all axes have

256 slices), z_{ij} represents the number of black slices along ith axis for the jth sample and k is a constant. The parameter k determines the number of 2D images in the dataset and is directly proportional to number of images ($\lfloor \ \rfloor$ represents the floor function).

3. We have introduced some rotational (as well as translational) independence in the model by arbitrarily rotating or flipping some images. This is illustrated in Fig. 1e–g.

4. We have also introduced slight normal, horizontal, and vertical Gaussian Blur in certain images (Fig. 1g–i) to reproduce the situation where the MRI may get blurred due to motion of the fetus at the time of scanning. This will help the model to learn some motion-correction properties during training. However, we have not applied this effect on the images which were derived from blurred 3D samples, i.e., they are blurred by default (Fig. 1c, for example).

5. Since we examined 2D slices, predicting the type and age from fetal brain is adequately possible only from the middle portion of the brain along any axis, which exhibits the overall development and the actual size of the brain and the brain parts, and thus precisely reflects

the age. As we move toward the edges, the brain (and its parts) start shrinking, consequently accurate brain age prediction is not possible. Thus, the type of prediction task, a classification problem, returns three classes: 'Neurotypical', 'Pathological' or 'Type cannot be determined from the slice'. The model returns "'Type cannot be determined from the slice" for the slices that do not belong to the intermediate portion of the brain. Same holds for the age prediction task as well, which is a regression problem. Lastly, the age values are normalized before giving it to the model.

The pre-processing steps mentioned above were applied to both the input images and their corresponding segmented outputs, except for the blurring effect, which was applied to only the input images. Thus, the final dataset consists of all sorts of images: neurotypical and pathological fetal brain images, flipped images, blurred images, and brain images of different sizes (depending on the age of the fetus at the time of scanning).

Finally, the input to the models is the pre-processed images, and their corresponding outputs included: the segmented images, the type of the fetal brain (including type cannot be predicted), and the gestational age of the fetus (including age cannot be predicted).

2.2 Network architecture

A U-Net architecture (Ronneberger et al. 2015) contains two main parts: the encoder, which learns to encode the essential data from the input image, and the decoder, which learns to decode the encoded data to return the required segmentation. We have tested out several models, keeping U-Net as the base model, where all models have the following in common:

1. The first step is to encode the necessary information from the input for all tasks, for which, we have used a set of 'Encoder blocks' each of which have two convolutional layers and a max-pool layer. These blocks are arranged in a layer-wise manner, with the size of the input reducing after each step. The output from the last encoder block then goes to the lowest layer, which consists of two convolutional layers.

2. For the segmentation task, the output from the second convolutional layer at the lowest layer is sent to a set of 'Decoder blocks', which consist of a deconvolutional layer (for decoding the inputs from the previous layer), a concatenation layer, and followed by two convolutional layers. These blocks are also arranged in a layer-wise manner, with the image size increasing after each step. The number of decoder blocks is the same as that of the encoder blocks. Also, an encoder block on each layer is connected to its corresponding decoder block in the same layer (the outputs of which are concatenated by the concatenation layer), either directly (as in the original U-Net) or via a convolutional layer.

3. For the type and age prediction task, the output from the first convolutional layer in the lowest layer is sent to a set of three dense layers which will encode the input further and is common for both the tasks. This output is then divided into two parts: one for type prediction, and other for age prediction; both tasks, thereby, having their own set of dense layers. The type prediction part will have two more dense layers and the age prediction part will have three dense layers, followed by their corresponding output layer (different combinations of dense layers have been tested out and the one mentioned here returned the best results).

4. The models take an input image of shape 256×256. The segmentation output of the shape $256 \times 256 \times 8$, as an eight-class classification (the seven brain components and the black background pixel) is performed for each pixel. The class with the highest probability for each pixel is then chosen as output. For type prediction, the output is of the shape 1×3, for three classes, and the one with the highest probability is selected as the output. For age prediction, the output is a single, normalized value, which is used to get the original age. A threshold for the normalized value is set and if the value is higher, then the original age is returned, otherwise, 'Age cannot be determined from the slice' is returned.

5. For the segmentation and type prediction task, sparse categorical cross entropy was used as the loss function and accuracy was used as the metric to train our model, and for the age prediction task, mean squared error was used both as the loss function and the metric. Adam's optimizer was chosen for training the models because they use fewer parameters than other optimizers (thus are faster), at the same time being efficient. Thirty epochs have been used to train the models, along with 'Early Stopping' and patience $= 3$ as the callback to prevent overfitting.

The lowest layer stores all the necessary encoded information about the input (and acts as a bottleneck); thus, its output is the most suitable input for the dense layer for further encoding, as compared to any other layer. The advantage of a combined model is that this model is capable of using the same encoding units and the same encoded information for all the tasks. They differ only in the second part: for the segmentation task where the encoded information is decoded using transposed convolution and for the other two tasks, it is passed on to dense layers for further encoding.

The models we have tried out, keeping the above specifications in common, are:

1. The first two models have five and six layers of encoder–decoder blocks, with the encoder and decoder blocks connected directly. The rest remain the same. The two models are named '5 layer no-mod multi-pred' and '6 layer no-mod multi-pred', respectively.

2. The next three models have five, six, and seven layers of encoder–decoder blocks, with the two blocks connected by a convolutional layer in between. This layer is added to be able to store additional information for segmentation; thus, the performance is expected to improve. These models are named '5-layer mod multi-pred', '6-layer mod multi-pred', and '7-layer mod multi-pred', respectively.

3. Next three models are also similar to the three models mentioned above. However, each one has two additional skip connections for the two prediction tasks. The skip connections will help eliminate the vanishing gradient problem, if any. These models are named '5-layer mod multi-pred skip', '6-layer mod multi-pred skip' (this is our selected model and is named 'MTSE U-Net', based on its characteristics), and '7-layer mod multi-pred skip', respectively.

We have trained multiple models with similar configurations as a part of our ablation study. This study will help us analyze the effect and importance of additional layers, skip connections, and the encoder–decoder blocks. The name of all the models used in this paper are highly descriptive of their architecture.

The architecture of the MTSE U-Net model is shown in Fig. 2, along with the components of encoder and decoder blocks. The values on each box represent the number of filters in case of encoder–decoder blocks, and convolutional layers, and the number of neurons in case of dense layers.

3 Results and discussion

After exploring the various models, they were tested out on the test set and on some individual images as well, to judge their performance results on various factors.

3.1 Evaluation metrics

The evaluation metrics used for the segmentation are: precision (P), sensitivity (S), Jaccard similarity (J), Dice score (D), and accuracy (A), as in Roy and Shoghi (2019); Rampun et al. 2019). Precision gives us a count of the number of correct predictions with respect to the ground truth values for any particular class, whereas sensitivity gives us a count of the number of correct predictions, with respect to all the predictions for any particular class. Jaccard similarity and
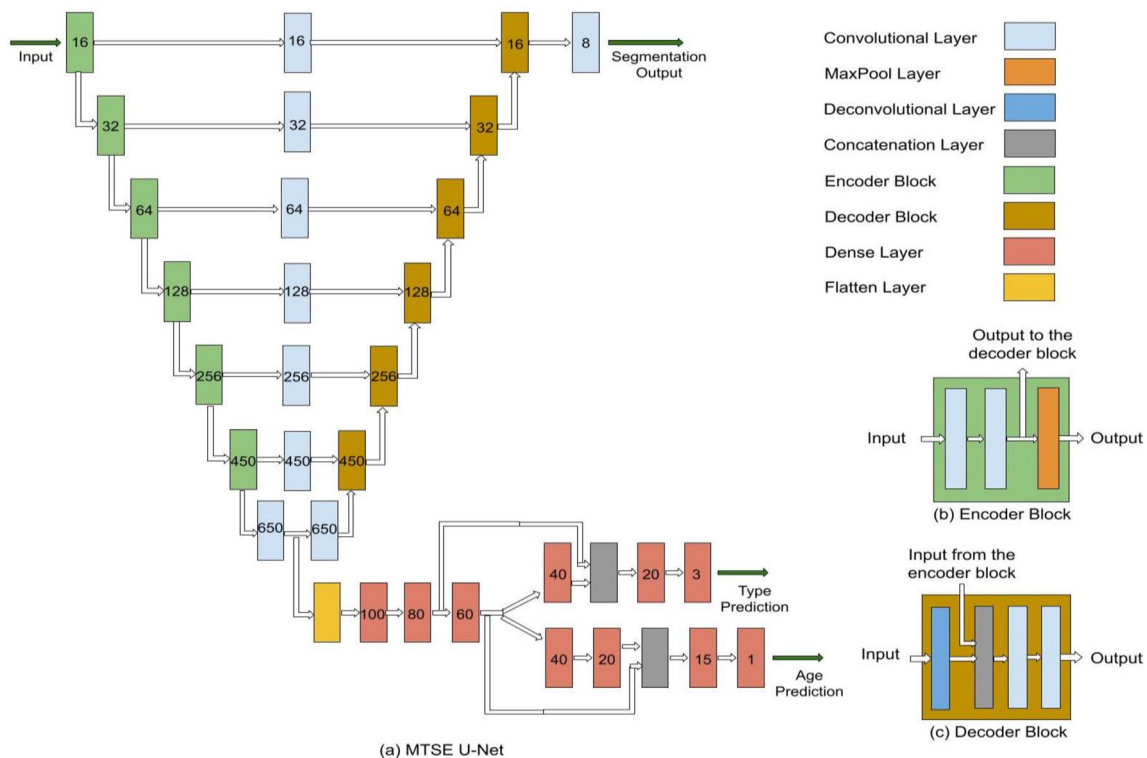

Fig. 2 Architecture of **a** MTSE U-Net, **b** encoder block, and **c** decoder block

Dice score give us the extent of overlapping of the prediction with the ground truth segmentation for each class. Accuracy is the total number of correctly predicted labels (pixels) in an image, irrespective of the class, by the total number of pixels in the image. The first four metrics are individually calculated for each class in every image. However, for some testing processes, we have used the average of these metrics over all the classes for any image:

$$\text{Average precision (AP)} = \frac{\sum_{\text{All classes}} P}{\text{Total number of classes}} \qquad (1)$$

$$\text{Average sensitivity (AS)} = \frac{\sum_{\text{All classes}} S}{\text{Total number of classes}} \qquad (2)$$

$$\text{Average Jaccard similarity (AJ)} = \frac{\sum_{\text{All classes}} J}{\text{Total number of classes}} \qquad (3)$$

$$\text{Average Dice score (AD)} = \frac{\sum_{\text{All classes}} D}{\text{Total number of classes}} \qquad (4)$$

Evaluation metric used for the fetal brain type prediction is accuracy (AT), which is the number of correct fetal brain type predictions from the brain images, with respect to the total number of images; and that for the gestational age prediction is mean absolute error (MAE), which is the difference between the actual and the predicted age for all the brain images, by the total number of images, and is calculated in weeks. The target is to maximize the metric values for the segmentation and type prediction task, and to minimize the MAE for the age prediction task.

3.2 Performance results

The test set of 550 images, consisting of all possible types of MRI slices, has been used to evaluate the performance of all the models, and compare them with the previously proposed works. Table 1 shows the detailed performance of each model, including the previous models (indicated in the row) on the whole test set for the segmentation of each of the seven brain part (indicated in the column), on the following metrics (and in the following order): *precision, sensitivity, Jaccard similarity,* and *Dice score*; all of which are represented in a single cell. It is expressed in the form: Mean ± SD, where 'Mean' and 'SD', respectively, represent the mean and the standard deviation of the performance of a model for the segmentation of a brain part. The model giving the highest value for each brain component and for each metric is indicated in bold. The highest values for each metric among the models developed in this work are underlined.

Table 1 shows that the models (developed in our work) giving the best segmentation performances are: MTSE U-Net (i.e., 6-layer mod multi-pred skip), followed by 5-layer mod multi-pred skip and 6-layer mod multi-pred; best results for most metrics are returned by the MTSE U-Net model. However, taking the previous works into consideration, it is seen that the best performance is returned by the U-Net architecture, having slightly better performance than that of our model, whereas for the other two models (i.e., Deep U-Net and FCN), it is somewhat similar to ours.

After the detailed performance on segmentation, the same test set has been used to evaluate the average performance of the models for all the three tasks (Table 2). The first five columns show the average evaluation metrics for the segmentation task, and the last two columns show the accuracy for type prediction task and MAE for the age prediction task, respectively. The metrics are expressed in the following form (for the segmentation and age prediction task only): 'Mean ± SD', where 'Mean' and 'SD' are the mean and standard deviation of the performance of the model (mentioned in the row) evaluated on the metric (mentioned in the column) for all the images in the test set. The accuracy of the type prediction task is a single value, representing its performance over the entire test set. For each metric, the model giving the highest value is indicated in bold. Also, the highest values for each metric among the models developed in this work are underlined.

It can again be seen that the best performing model for the segmentation task is U-Net, having slightly better performance than that of MTSE U-Net, and that of Deep U-Net and FCN is again very much similar to our MTSE U-Net model. The lowest value for the age prediction task is returned by the 5-layer mod multi-pred model (MAE = 0.82 weeks). However, when all the three tasks are considered, the overall best performing combined task model is MTSE U-Net, returning great results for all metrics. The segmentation performance of the MTSE U-Net model exceeds that of other combined task models by approximately 2–3 percent. It also returned the best type prediction accuracy (which is also returned by the 5-layer mod multi-pred skip model). We have also plotted a visualization of the average segmentation results of MTSE U-Net on the test set in Fig. 3. The bar represents the mean of the specified metric and the vertical line represents the standard deviation with respect to the mean.

Table 2 shows that we did not compare our type and age prediction performance with the previously proposed methods. This is because, as per our knowledge, no previous work on fetal brain classification into neurotypical and pathological type exists; thus, the performance comparison for this task is restricted to our models only. And the previous works on gestational age prediction from 2D brain images (Shi et al. 2020; Shen et al. 2022) did not consider the possibility

Table 1 Evaluation of all the models on the four evaluation metrics (precision, sensitivity, Jaccard similarity, and Dice score) for only the segmentation task of each individual brain part

	CSF	GM	WM	LV	CBM	SGM	BS
Deep U-Net	0.80 ± 0.26	**0.74 ± 0.22**	0.90 ± 0.15	0.88 ± 0.25	0.86 ± 0.31	0.83 ± 0.30	0.82 ± 0.34
	0.72 ± 0.31	0.71 ± 0.25	0.89 ± 0.17	0.85 ± 0.23	0.94 ± 0.19	**0.93 ± 0.16**	0.87 ± 0.29
	0.62 ± 0.31	0.61 ± 0.28	0.81 ± 0.20	0.77 ± 0.30	0.82 ± 0.34	0.79 ± 0.32	0.74 ± 0.39
	0.71 ± 0.32	0.71 ± 0.25	**0.88 ± 0.18**	0.83 ± 0.28	0.84 ± 0.33	0.83 ± 0.30	0.77 ± 0.38
FCN	0.75 ± 0.29	0.71 ± 0.25	0.86 ± 0.22	0.87 ± 0.24	**0.91 ± 0.25**	0.84 ± 0.30	**0.88 ± 0.28**
	0.75 ± 0.29	0.74 ± 0.22	0.90 ± 0.16	0.87 ± 0.21	0.90 ± 0.24	0.91 ± 0.23	0.84 ± 0.32
	0.62 ± 0.31	0.59 ± 0.28	0.79 ± 0.24	0.78 ± 0.29	**0.84 ± 0.29**	0.78 ± 0.34	0.79 ± 0.36
	0.70 ± 0.32	0.70 ± 0.25	0.85 ± 0.23	0.83 ± 0.27	**0.86 ± 0.27**	0.82 ± 0.32	0.82 ± 0.34
U-Net	**0.82 ± 0.23**	0.73 ± 0.21	**0.90 ± 0.16**	**0.88 ± 0.18**	0.87 ± 0.30	**0.90 ± 0.22**	0.87 ± 0.25
	0.74 ± 0.31	**0.75 ± 0.20**	0.90 ± 0.14	0.88 ± 0.19	**0.94 ± 0.14**	0.91 ± 0.18	**0.89 ± 0.23**
	0.62 ± 0.31	**0.63 ± 0.26**	**0.82 ± 0.19**	0.80 ± 0.24	0.83 ± 0.32	**0.83 ± 0.26**	**0.81 ± 0.31**
	0.71 ± 0.31	**0.73 ± 0.31**	0.87 ± 0.17	**0.85 ± 0.21**	0.85 ± 0.31	**0.88 ± 0.24**	**0.85 ± 0.29**
5 Layer no-mod multi-pred	0.62 ± 0.34	0.69 ± 0.29	0.77 ± 0.28	0.80 ± 0.32	0.74 ± 0.40	0.78 ± 0.36	0.80 ± 0.36
	<u>**0.80 ± 0.26**</u>	0.64 ± 0.27	0.92 ± 0.13	0.86 ± 0.24	0.87 ± 0.29	0.88 ± 0.25	0.82 ± 0.34
	0.55 ± 0.32	0.50 ± 0.30	0.72 ± 0.28	0.72 ± 0.34	0.67 ± 0.43	0.71 ± 0.39	0.68 ± 0.42
	0.64 ± 0.34	0.61 ± 0.29	0.79 ± 0.27	0.77 ± 0.32	0.70 ± 0.43	0.74 ± 0.37	0.71 ± 0.41
6 Layer no-mod multi-pred	0.75 ± 0.29	0.73 ± 0.25	0.80 ± 0.25	0.86 ± 0.25	0.80 ± 0.36	0.84 ± 0.31	0.85 ± 0.31
	0.73 ± 0.29	0.67 ± 0.26	0.92 ± 0.13	0.84 ± 0.25	<u>0.90 ± 0.25</u>	0.88 ± 0.25	0.84 ± 0.32
	0.61 ± 0.31	0.55 ± 0.30	0.75 ± 0.26	0.76 ± 0.31	0.74 ± 0.39	0.77 ± 0.35	0.74 ± 0.39
	0.69 ± 0.32	0.66 ± 0.27	0.82 ± 0.25	0.81 ± 0.28	0.76 ± 0.39	0.80 ± 0.33	0.76 ± 0.38
5 Layer mod multi-pred	0.72 ± 0.32	0.73 ± 0.25	0.79 ± 0.27	0.82 ± 0.29	0.83 ± 0.34	0.81 ± 0.34	0.85 ± 0.32
	0.75 ± 0.28	0.67 ± 0.26	0.92 ± 0.13	0.86 ± 0.23	0.89 ± 0.26	0.87 ± 0.27	<u>0.86 ± 0.30</u>
	0.60 ± 0.32	0.55 ± 0.30	0.74 ± 0.28	0.73 ± 0.33	0.75 ± 0.39	0.72 ± 0.38	0.74 ± 0.38
	0.69 ± 0.33	0.66 ± 0.27	0.81 ± 0.26	0.79 ± 0.31	0.77 ± 0.38	0.75 ± 0.37	0.77 ± 0.37
6 Layer mod multi-pred	<u>0.77 ± 0.28</u>	**0.74 ± 0.25**	0.80 ± 0.25	0.87 ± 0.26	0.86 ± 0.31	<u>**0.90 ± 0.24**</u>	<u>**0.88 ± 0.29**</u>
	0.73 ± 0.30	0.67 ± 0.26	**0.94 ± 0.12**	0.84 ± 0.24	<u>0.90 ± 0.24</u>	0.85 ± 0.28	0.85 ± 0.31
	0.62 ± 0.31	0.56 ± 0.29	0.76 ± 0.26	0.75 ± 0.31	0.79 ± 0.36	<u>0.79 ± 0.34</u>	0.76 ± 0.37
	0.70 ± 0.32	0.67 ± 0.27	<u>0.84 ± 0.25</u>	0.81 ± 0.28	0.81 ± 0.35	0.82 ± 0.32	0.79 ± 0.35
7 Layer mod multi-pred	0.73 ± 0.29	<u>**0.74 ± 0.25**</u>	0.83 ± 0.24	0.85 ± 0.26	0.88 ± 0.30	0.85 ± 0.30	0.82 ± 0.35
	0.78 ± 0.28	0.67 ± 0.26	0.91 ± 0.13	0.87 ± 0.22	0.88 ± 0.26	0.88 ± 0.27	0.85 ± 0.29
	0.62 ± 0.31	0.56 ± 0.29	0.76 ± 0.25	0.76 ± 0.30	0.79 ± 0.36	0.77 ± 0.35	0.72 ± 0.39
	<u>**0.71 ± 0.31**</u>	0.67 ± 0.26	<u>0.84 ± 0.24</u>	0.82 ± 0.28	0.81 ± 0.35	0.81 ± 0.34	0.75 ± 0.38
5 Layer mod multi-pred skip	0.75 ± 0.29	0.73 ± 0.24	0.84 ± 0.23	**0.88 ± 0.23**	<u>**0.91 ± 0.26**</u>	0.82 ± 0.32	0.83 ± 0.33
	0.76 ± 0.28	0.72 ± 0.23	0.90 ± 0.14	0.85 ± 0.24	<u>0.90 ± 0.26</u>	<u>0.91 ± 0.22</u>	0.86 ± 0.29
	<u>**0.63 ± 0.31**</u>	0.58 ± 0.28	<u>0.78 ± 0.25</u>	<u>0.77 ± 0.30</u>	0.82 ± 0.34	0.76 ± 0.36	0.74 ± 0.38
	<u>**0.71 ± 0.31**</u>	0.69 ± 0.25	<u>0.84 ± 0.23</u>	0.82 ± 0.27	0.84 ± 0.33	0.80 ± 0.34	0.77 ± 0.37
6 Layer mod multi-pred skip	0.75 ± 0.29	0.72 ± 0.25	<u>0.86 ± 0.23</u>	0.85 ± 0.26	<u>**0.91 ± 0.25**</u>	0.86 ± 0.28	<u>**0.88 ± 0.28**</u>
	0.77 ± 0.28	**0.75 ± 0.22**	0.89 ± 0.15	**0.89 ± 0.21**	<u>0.90 ± 0.24</u>	0.90 ± 0.24	0.85 ± 0.30
	<u>**0.63 ± 0.31**</u>	<u>0.60 ± 0.28</u>	0.78 ± 0.28	<u>0.77 ± 0.30</u>	<u>**0.84 ± 0.32**</u>	<u>0.79 ± 0.33</u>	<u>0.77 ± 0.37</u>
	<u>**0.71 ± 0.32**</u>	<u>0.71 ± 0.25</u>	<u>0.84 ± 0.23</u>	<u>0.83 ± 0.28</u>	<u>**0.86 ± 0.30**</u>	<u>0.83 ± 0.31</u>	<u>0.80 ± 0.35</u>
7 Layer mod multi-pred skip	<u>0.77 ± 0.29</u>	0.72 ± 0.26	0.78 ± 0.27	0.85 ± 0.28	0.82 ± 0.34	0.85 ± 0.30	0.84 ± 0.32
	0.71 ± 0.31	0.67 ± 0.27	0.92 ± 0.12	0.84 ± 0.24	<u>0.90 ± 0.26</u>	0.86 ± 0.28	0.85 ± 0.31
	0.60 ± 0.32	0.54 ± 0.30	0.74 ± 0.27	0.74 ± 0.32	0.75 ± 0.39	0.76 ± 0.36	0.73 ± 0.39
	0.68 ± 0.33	0.65 ± 0.28	0.81 ± 0.26	0.80 ± 0.30	0.77 ± 0.38	0.80 ± 0.34	0.76 ± 0.37

The model giving the highest value for each brain component and for each metric is indicated in bold. The highest values for each metric among the models developed in this work are underlined

of a slice being away from the center of the brain. However, as a segmentation input can be from any portion of the brain, we had to train our model accordingly (by adding a new category for the non-central slices), and thus our processes are not directly comparable.

After identifying MTSE U-Net as the best model, we have also performed threefold cross-validation on this model to confirm that it is robust and is not biased by the training data. The results are shown in Table 3. The three iterations in this table indicate the three different combinations of train and test set for the threefold cross-validation.

As can be seen from the above table, the results across all the iterations are very similar, showing that our model is consistent in its results and robust in any situation. It gave good results on a test set of more than 4,700 samples. However, it is seen that these results are a bit lower that that seen

Table 2 Evaluation of the models for all the three tasks on the five average metrics for the segmentation task, accuracy for the type prediction task, and MAE for the age prediction task

	AP	AS	AJ	AD	A	AT	MAE
Deep U-Net	0.85 ± 0.13	**0.87 ± 0.11**	0.77 ± 0.17	0.82 ± 0.15	**0.97 ± 0.02**	–	–
FCN	0.86 ± 0.12	0.86 ± 0.11	0.77 ± 0.16	0.82 ± 0.14	**0.97 ± 0.02**	–	–
U-Net	**0.87 ± 0.11**	**0.87 ± 0.09**	**0.79 ± 0.14**	**0.85 ± 0.12**	**0.97 ± 0.02**	–	–
5-Layer no-mod multi-pred	0.78 ± 0.15	0.85 ± 0.13	0.70 ± 0.18	0.75 ± 0.16	0.96 ± 0.03	0.87	1.01 ± 2.07
6-Layer no-mod multi-pred	0.83 ± 0.13	0.85 ± 0.13	0.74 ± 0.18	0.79 ± 0.15	**$\underline{0.97 \pm 0.03}$**	0.88	1.07 ± 2.42
5-Layer mod multi-pred	0.82 ± 0.13	0.85 ± 0.12	0.73 ± 0.18	0.78 ± 0.16	**$\underline{0.97 \pm 0.03}$**	0.87	**$\underline{0.82 \pm 1.88}$**
6-Layer mod multi-pred	$\underline{0.85 \pm 0.13}$	0.85 ± 0.12	0.75 ± 0.17	0.80 ± 0.15	**$\underline{0.97 \pm 0.03}$**	0.88	0.94 ± 2.16
7-Layer mod multi-pred	0.84 ± 0.14	0.85 ± 0.12	0.75 ± 0.18	0.80 ± 0.16	**$\underline{0.97 \pm 0.03}$**	0.78	0.87 ± 1.65
5-Layer mod multi-pred skip	0.84 ± 0.13	0.86 ± 0.12	$\underline{0.77 \pm 0.17}$	0.81 ± 0.15	**$\underline{0.97 \pm 0.03}$**	$\underline{0.89}$	0.99 ± 2.28
6-Layer mod multi-pred skip	$\underline{0.85 \pm 0.12}$	$\underline{0.87 \pm 0.11}$	$\underline{0.77 \pm 0.16}$	$\underline{0.82 \pm 0.14}$	**$\underline{0.97 \pm 0.02}$**	$\underline{0.89}$	0.83 ± 2.02
7-Layer mod multi-pred skip	0.83 ± 0.14	0.84 ± 0.13	0.73 ± 0.19	0.79 ± 0.17	**$\underline{0.97 \pm 0.03}$**	0.84	0.98 ± 1.84

The model giving the highest value for each brain component and for each metric is indicated in bold. The highest values for each metric among the models developed in this work are underlined
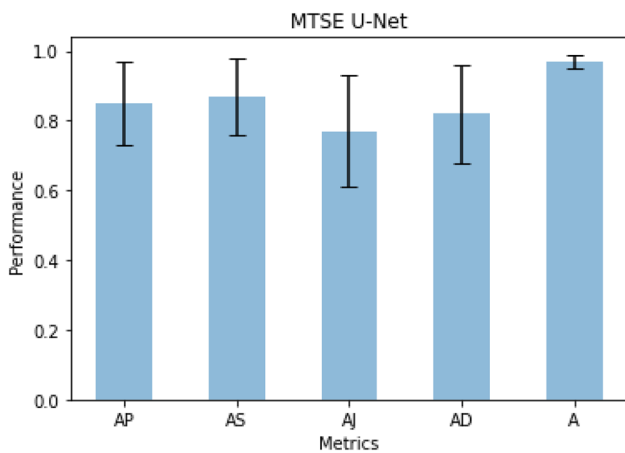
Fig. 3 Bar plot showing mean and SD on various segmentation metrics of MTSE U-Net

in Table 2. This can be attributed to the fact that threefold cross-validation requires dividing the entire dataset into three equal parts. Thus, there were around 9,500 samples in the training dataset for cross-validation (which is 4,000 data less than our actual training set).

3.3 Discussion

The three tasks that we are trying to solve in this work are different, even though they require the same input; as

a result, their encoded information is also somewhat different (for example, the segmentation task would require the spatial information, which the other two tasks might remove). Thus, our combined task model parameters, especially the encoding unit, has to be fine tuned to store the necessary encodings for all the tasks. This gives rise to four important observations:

1. Tables 1 and 2 show that the combined task models having a convolutional layer in between the encoder and the decoder blocks gave much better performance than a direct connection between the blocks. Thus, these additional convolutional layers store those attributes/ features, which are essential only for the segmentation purpose.

2. Upon comparing our combined task model with the trained, solely for performing just one task (segmentation, for example), it is possible that the performance of the latter model might be better than ours for the same task; as was the case with the U-Net model (even though the performance difference is very small), where the encoder was required to store parameters for the segmentation task only. However, the performance of our model seems to be very impressive, when the performances on all the three tasks are taken into consideration, given that no individual model was created for these tasks.

Table 3 Evaluation results of MTSE U-Net model on threefold cross-validation

	AP	AS	AJ	AD	A	AT	MAE
First iteration	0.81 ± 0.16	0.85 ± 0.13	0.75 ± 0.20	0.78 ± 0.17	0.97 ± 0.03	0.85	0.92 ± 1.89
Second iteration	0.83 ± 0.15	0.86 ± 0.11	0.73 ± 0.18	0.78 ± 0.16	0.97 ± 0.03	0.88	0.94 ± 2.03
Third iteration	0.84 ± 0.13	0.85 ± 0.12	0.74 ± 0.17	0.80 ± 0.15	0.97 ± 0.02	0.86	1.07 ± 2.38

3. A trade-off situation may possibly occur in between the performances of the segmentation and the prediction tasks due to the differences in the required encoding for these tasks. The trade-off is highest in the 'no-mod multi-pred' models (Table 2). This effect has been reduced to some extent in the 'mod multi-pred' models, including the ones using a skip connection, because of the additional convolutional layer. It is also possible that the task 'dominating' the encoding unit may have an advantage over the other two.

4. Using a common encoder for all the three tasks has an advantage of reduced space required to store the model parameters, as compared to the combined space required for storing the model parameters performing these tasks individually. Less parameters imply less time taken by our model to give an output, and also less computational power requirement. Thus, a single model for the three tasks not only allows us to perform them simultaneously, but also ends up saving space, time, and energy.

Using a skip connection for the age and type prediction tasks has been found to be beneficial. The models using skip connections returned better results for both the segmentation and the prediction tasks, than their counterpart models not using one. Furthermore, the MTSE U-Net model showed the minimum above-mentioned trade-off, thus giving satisfactory performances for all the three tasks.

Table 1 also shows that increasing the number of layers does not necessarily improve the performance of the model as in the case with the two 7-layer models, possibly because of the fact that excessive reduction in the image size also results in the loss of necessary information.

After checking the performance on the test set, we have compared the performance of the 5-layer mod multi-pred skip, 6-layer mod multi-pred, and MTSE U-Net models on two images to compare the quality of the outputs they return in Table 4. It shows the input image, the actual segmentation, the outputs returned by each model on all the three tasks and their evaluation results.

Both images are normal and non-blurred. For the first image, all the models returned accurate results, (though the second model returned slightly lower segmentation results than the other two). For the second image, the first model gave the worst results, particularly the wrong type prediction and a high MAE for age prediction. The next two models returned highly accurate predictions.

The MTSE U-Net model is tested on five more images of varying types to evaluate its performance. The images are arranged in a row-wise manner, with the actual segmentation, the model's output, and the evaluation result shown just beside it on the same row in Fig. 4.

The model has been quite accurate with the segmentation of the first four images, of which, second and fourth are normal images and the other two are blurred images. The type prediction for all the images is correct. Type and age cannot be predicted for the first and the fourth image, as these slices do not belong to the middle portion of the brain. The age prediction for the second and third images is highly accurate. Finally, it can be seen that the performance of the model on the last image is not that good; however, upon observing the segmentation output, it is noticeable that the model has tried to smooth out the blurred portion of the input image (which is an effect of the motion correction we have introduce in the model), whereas the actual segmentation has maintained this effect. This difference can only be seen in those images, which were blurred by default in the dataset (and not in the ones, which we blurred out in the preprocessing step), and may account for the reduced segmentation metric values to some extent. Thus, the model's performance on blurred images is also pretty impressive.
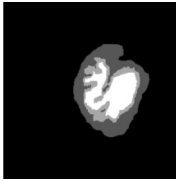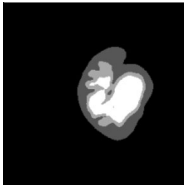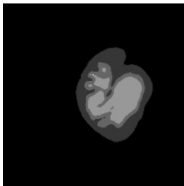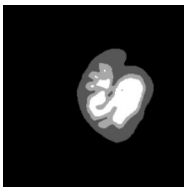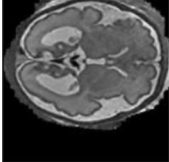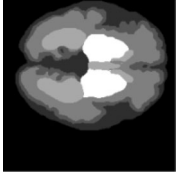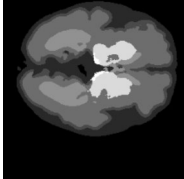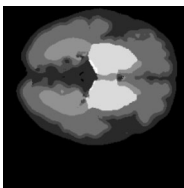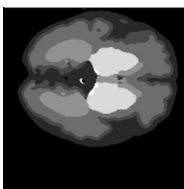
Despite low MAE, the age prediction task (for very few cases) is somewhat prone to relatively large errors in predictions, which accounts for the high standard deviation. This may be attributed to three reasons:

1. A 2D slice slightly away from the center of the brain may somewhat resemble the center slice of a younger fetal brain. Thus, the model may confuse between such slices.

2. Although the model was trained to predict age for both brain types, yet, a pathological brain highly differs from a neurotypical brain in its brain structure (Fig. 5), which might confuse the model into making wrong predictions. Also, different pathological conditions give rise to different fetal brain structures. Thus, an ambiguity in age prediction remains.

3. Even though the model was also trained to predict age from blurred images, yet, highly blurred out images often obscure the overall development in the brain (as was the case with the last image in Fig. 4); thus, the model may produce wrong predictions in such cases.

Finally, we have also shown the salient regions identified by the bottom-most convolutional layer of the model for type and age prediction in four input images (Fig. 6). We have used Grad-CAM (Selvaraju et al. 2017) for this purpose and these regions are represented as heat maps overlaid on the actual image. The lowest layer is chosen because it will best represent the features learnt/extracted by it or any of its previous layers; incorrect region identification represents a wrong feature extraction by this layer, or any of its previous layers.

The red regions in Fig. 6 represent the regions that (as per the model) are most influential in making the prediction, followed by the yellow, green, and blue regions (which represent the least significant regions). It can be seen that

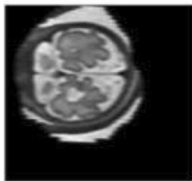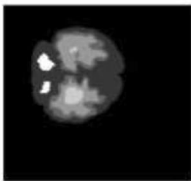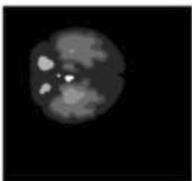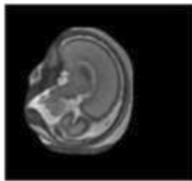**Table 4** Testing the performance of three models on two images
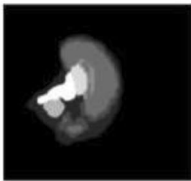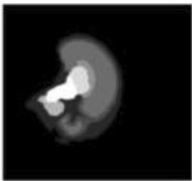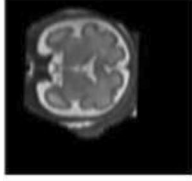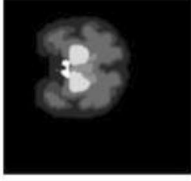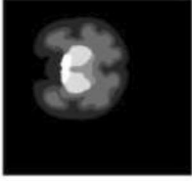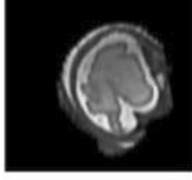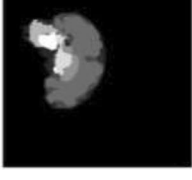
Input image	Actual segmentation	Model used	Outputs			Evaluation
			Segmentation	Type prediction	Age prediction	
		6 Layer mod multi-pred		'Type cannot be determined from the slice'	'Age cannot be determined from the slice'	AP = 0.94 AS = 0.93 AJ = 0.90 AD = 0.94 A = 0.97 AT = 1.0 MSE = 0.0
		5 Layer mod multi-pred skip		'Type cannot be determined from the slice	'Age cannot be determined from the slice'	AP = 0.83 AS = 0.94 AJ = 0.78 AD = 0.82 A = 0.98 AT = 1.0 MSE = 0.0
		MTSE U-Net		'Type cannot be determined from the slice'	'Age cannot be determined from the slice'	AP = 0.95 AS = 0.94 AJ = 0.90 AD = 0.94 A = 0.97 AT = 1.0 MSE = 0.0
		6 Layer mod multi-pred		'Neurotypical'	29.6 weeks	AP = 0.80 AS = 0.86 AJ = 0.67 AD = 0.76 A = 0.91 AT = 0.0 MSE = 3.5
		5 Layer mod multi-pred skip		'Pathological'	33.0 weeks	AP = 0.80 AS = 0.91 AJ = 0.73 AD = 0.79 A = 0.93 AT = 1.0 MSE = 0.1
		MTSE U-Net		'Pathological'	33.1 weeks	AP = 0.80 AS = 0.91 AJ = 0.73 AD = 0.79 A = 0.93 AT = 1.0 MSE = 0.0

for the first three images, the layer identified the regions accurately for proceeding with its predictions (the model also captured a small background region close to the brain with a low confidence, as indicated in the heat map. This is required to make sure that the model did not miss out on any of the brain part in the image while making predictions). But the model identified a wrong region in the last image. However, such cases are very few, as we found during experimenting with different input images and also as indicated by the prediction results. Finally, these region highlights are not required for the segmentation task because

the task performance itself represents the model's salient region identification capability.

Now, as stated earlier, all of the models were created using 2D slices of the 3D samples. Thus, a possible future work can be to extend this work to the third dimension. This might help, especially, in the type and age prediction task, and also provide a remedy to the above-mentioned problem. For example, the model can more surely conclude that the age cannot be predicted from a group of slices if it sees a continuous increase or decrease in the size of the brain, indicating that the center of the brain is

Fig. 4 Evaluating the MTSE U-Net model on five images

| Brain Image	Actual Segmentation	Segmentation Output	Type and Age Predictions	Evaluation Metric values
			'Type cannot be determined from the slice' 'Age cannot be determined from the slice'	AP = 0.78 AS = 0.89 AJ = 0.70 AD = 0.77 A = 0.97 AT = 1.0 MAE = 0.0
			'Pathological' 27.4 weeks	AP = 0.83 AS = 0.83 AJ = 0.72 AD = 0.83 A = 0.97 AT = 1.0 MAE = 0.8
			'Neurotypical' 30.6 weeks	AP = 0.85 AS = 0.90 AJ = 0.77 AD = 0.86 A = 0.96 AT = 1.0 MAE = 0.3
			'Type cannot be determined from the slice' 'Age cannot be determined from the slice'	AP = 0.95 AS = 0.95 AJ = 0.91 AD = 0.95 A = 0.97 AT = 1.0 MAE = 0.0
			'Pathological' 26.9 weeks	AP = 0.60 AS = 0.73 AJ = 0.50 AD = 0.64 A = 0.93 AT = 1.0 MAE = 4.3

Fig. 5 Difference in the fetal brain structures between neurotypical and pathological fetal brains of similar gestational age

Neurotypical Brain Age: 30.3 weeks   Neurotypical Brain Age: 30 weeks   Pathological Brain Age: 30.4 weeks
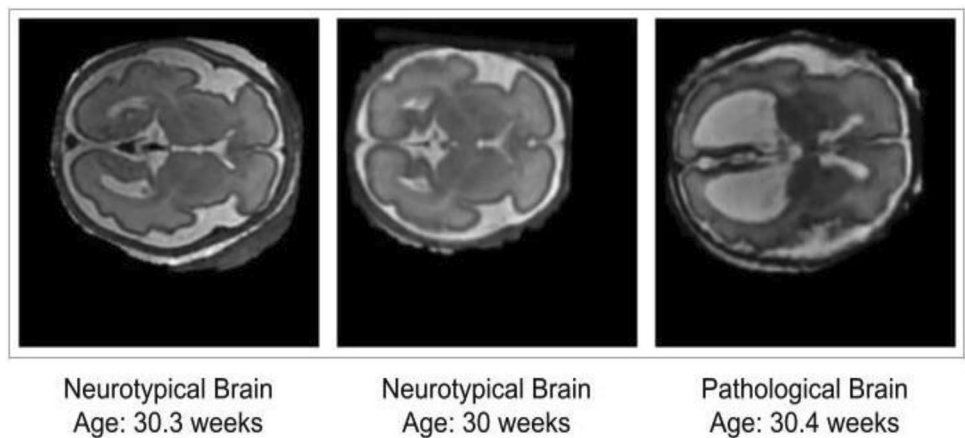
not present in the group of slices for the model to make accurate predictions. This facility cannot be utilized in 2D slices. Other future prospects can include classification of the fetal brain into more specific conditions like Dandy–Walker malformation, colpocephaly, mega-cisterna manga, etc., and search for additional features and observations and develop models that work on 2D slices and provide better performance.

Fig. 6 Heat map representing the salient regions in the input image for age and type prediction
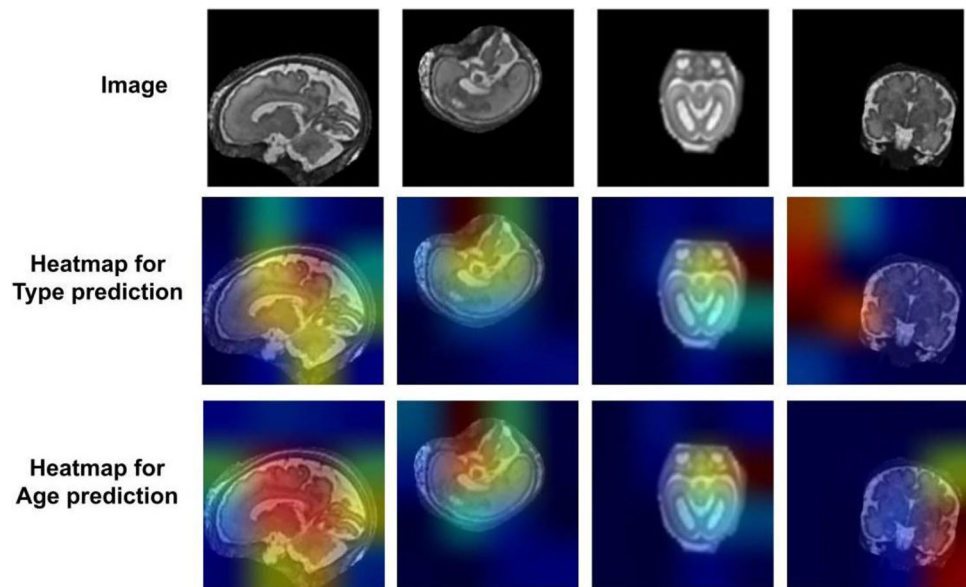
4 Conclusion

In this paper, we proposed MTSE U-Net, a deep learning architecture, for performing three tasks (and thus having three outputs): segmentation of the brain into its seven major components and prediction of the fetal brain type ('Pathological' or 'Neurotypical') and the gestational age at the time of scanning. The input is the 2D fetal brain slices. This can be obtained after segmenting the fetal MRI to return the fetal brain image, which explains our work as an extension of already existing works on segmentation of the fetal MRI. Our work is inspired by the U-Net architecture. We have tested out the model on our test set and also on some individual inputs and analyzed them. The results for all the tasks have been convincing. We have also compared our model with some state-of-the-art models and showed that it can perform multiple tasks simultaneously with good accuracy, thus eliminating the need to use individual models for all of these tasks. The limitations and drawbacks of the models have also been discussed above along with the possible future works.

Declarations

References

Chi JG, Dooling EC, Gilles FH (1977) Gyral development of the human brain. Annals of Neurology. Wiley, New York, pp 86–93

Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3D U-Net: learning dense volumetric segmentation from sparse annotation. International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 424–432

Do N-T, Jung S-T, Yang H-J, Kim S-H (2021) Multi-Level Seg-Unet Model with Global and Patch-Based X-ray Images for Knee Bone Tumor Detection In Diagnostics. MDPI AG 11(4):691. https://doi.org/10.3390/diagnostics11040691

Gale CR (2004) Critical periods of brain growth and cognitive function in children. brain. Oxford University Press (OUP), Oxford, pp 321–329

Hagerty, Jason Stanley, Ronald Stoecker, William. (2017). Medical Image Processing in the Age of Deep Learning-Is There Still Room for Conventional Medical Image Processing Techniques? https://doi.org/10.5220/0006273803060311.

Hosny IA, Elghawabi HS (2010) Ultrafast MRI of the fetus: an increasingly important tool in prenatal diagnosis of congenital anomalies. Magn Reson Imaging 28(10):1431–1439. https://doi.org/10.1016/j.mri.2010.06.024 (**Epub 2010 Sep 17 PMID: 20850244**)

Jaimes C, Rofeberg V, Stopp C, Ortinau CM, Gholipour A, Friedman KG, Tworetzky W, Estroff J, Newburger JW, Wypij D, Warfield SK, Yang E, Rollins CK (2020) Association of Isolated Congenital Heart Disease with Fetal Brain Maturation. Am J Neuroradiol 41(8):1525–1531

Joseph R (2014) Brain tumor mri image segmentation and detection in image processing. Int Jurnal Res Eng Technol 03:1–5. https://doi.org/10.15623/ijret.2014.0313001

Khalili N, Lessmann N, Turk E, Claessens N, de Heus R, Kolk T, Viergever MA, Benders MJNL, Išgum I (2019) Automatic brain tissue segmentation in fetal MRI using convolutional neural networks. Magn Reson Imaging 64:77–89. https://doi.org/10.1016/j.mri.2019.05.020

Konkel L (2018) The brain before birth: using fMRI to explore the secrets of fetal neurodevelopment in environmental health perspectives. Environ Health Perspect 126:112001. https://doi.org/10.1289/ehp2268

Levine D (2001) Ultrasound versus magnetic resonance imaging in fetal evaluation. Topic Magn Resonan Imaging 12(1):25–38

Levman J, Takahashi E (2015) Multivariate analyses applied to fetal, neonatal and pediatric MRI of neurodevelopmental disorders. NeuroImage: Clinical. Elsevier BV, Amsterdam, pp 532–544

Liao L et al (2020) Multi-Branch Deformable Convolutional Neural Network with Label Distribution Learning for Fetal Brain Age Prediction: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). pp 424–427. doi: https://doi.org/10.1109/ISBI45749.2020.9098553

Lou J, Li D, Bui TD, Zhao F, Sun L, Li G, Shen D (2019) Automatic fetal brain extraction using multi-stage U-Net with deep supervision. Machine Learning in Medical Imaging. Springer International Publishing, Cham, pp 592–600

Murray RM, Jones P, O'Callaghan E (1991) Fetal brain development and later schizophrenia. Child Environ Adult Dis 156:155

Payette K, de Dumast P, Kebiri H et al (2021) An automatic multi-tissue human fetal brain segmentation benchmark using the fetal tissue annotation dataset. Sci Data 8:167. https://doi.org/10.1038/s41597-021-00946-3

Rajchl M, Lee MCH, Schrans F, Davidson A, PasseratPalmbach J, Tarroni G, Alansary A, Oktay O, Kainz B, Rueckert D (2016) Learning under distributed weak supervision (Version 1). Arxiv. https://doi.org/10.48550/ARXIV.1606.01100

Rampun A, Jarvis D, Griffiths P, Armitage P (2019) Automated 2D fetal brain segmentation of mr images using a deep u-net. In Asian Conference on Pattern Recognition. Springer, Cham 373–386

Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In international conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 234–241

Roy S, Bandyopadhyay SK (2016) A new method of brain tissues segmentation from MRI with accuracy estimation. Procedia Comput Sci 85:362–369

Roy S, Shoghi KI (2019) Computer-Aided Tumor Segmentation from T2-Weighted MR Images of Patient-Derived Tumor Xenografts. In: Karray F, Campilho A, Yu A (eds) Image Analysis and Recognition ICIAR 2019 Lecture Notes in Computer Science. Springer, Cham, pp 159–179

Roy S, Bhattacharyya D, Bandyopadhyay SK, Kim TH (2017a) An improved brain MR image binarization method as a preprocessing

for abnormality detection and features extraction. Front Comp Sci 11(4):717–727

Roy S, Bhattacharyya D, Bandyopadhyay SK, Kim TH (2017b) An iterative implementation of level set for precise segmentation of brain tissues and abnormality detection from MR images. IETE J Res 63(6):769–783

Roy S, Whitehead TD, Li S et al (2022) Co-clinical FDG-PET radiomic signature in predicting response to neoadjuvant chemotherapy in triple-negative breast cancer. Eur J Nucl Med Mol Imaging 49:550–562. https://doi.org/10.1007/s00259-021-05489-8

Salehi SSM et al (2018) Real-time automatic fetal brain extraction in fetal MRI by deep learning.IEEE 15 th International Symposium on Biomedical Imaging (ISBI 2018), pp 720–724. doi: https://doi.org/10.1109/ISBI.2018.8363675.

Scher MS (2003) Prenatal contributions to epilepsy: lessons from the bedside. Epileptic Disord 5(2):77–91 (**PMID: 12875951**)

Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: visual explanations from deep networks via gradient-based localization. IEEE Int Conf Comput vis (ICCV) 2017:618–626. https://doi.org/10.1109/ICCV.2017.74

Shen D, Wu G, Suk H-I (2017) Deep learning in medical image analysis. In annual review of biomedical engineering. Ann Rev 19:221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Shen L, Zheng J, Lee EH et al (2022) Attention-guided deep learning for gestational age prediction using fetal brain MRI. Sci Rep 12:1408. https://doi.org/10.1038/s41598-022-05468-5

Shi Y, Xue Y, Chen C, Lin K, Zhou Z (2020) Association of gestational age with MRI-based biometrics of brain development in fetuses. In BMC Medical Imaging. https://doi.org/10.1186/s12880-020-00525-9 (**Springer Science and Business Media LLC**)

Xu F, Ma H, Sun J,Wu R, Liu X, Kong Y (2019) LSTM Multi-modal UNet for Brain Tumor Segmentation. IEEE 4th International Conference on Image, Vision and Computing (ICIVC) pp 236-240. https://doi.org/10.1109/ICIVC47709.2019.8981027.

Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J (2018) UNet++: a nested U-Net architecture for medical image segmentation. In deep learning in medical image analysis and multimodal learning for clinical decision support. Springer International Publishing, Cham, pp 3–11