Recognition of Fetal Facial Expressions Using Artificial Intelligence Deep Learning

Yasunari Miyagi1, Toshiyuki Hata2, Saori Bouno3, Aya Koyanagi4, Takahito Miyake5

ABSTRACT

Fetal facial expressions are useful parameters for assessing brain function and development in the latter half of pregnancy. Previous investigations have studied subjective assessment of fetal facial expressions using four-dimensional ultrasound. Artificial intelligence (AI) can enable the objective assessment of fetal facial expressions. Artificial intelligence recognition of fetal facial expressions may open the door to the new scientific field, such as “AI science of fetal brain”, and fetal neurobehavioral science using AI is at the dawn of a new era. Our knowledge of fetal neurobehavior and neurodevelopment will be advanced through AI recognition of fetal facial expressions. Artificial intelligence may be an important modality in current and future research on fetal facial expressions and may assist in the evaluation of fetal brain function.

Keywords: Artificial intelligence, Deep learning, Facial recognition, Fetus, Machine learning, Ultrasonography.

Donald School Journal of Ultrasound in Obstetrics and Gynecology (2021): 10.5005/jp-journals-10009-1710

INTRODUCTION

Fetal behaviors such as fetal movements and facial expressions that have been observed by four- (4D) or three-dimensional (3D) ultrasound have been deemed to be related to the development of fetal central nervous system development.1–11 A scoring system,12 which was originally reported by Kurjak et al. and later modified by Stanojevic et al.,13 can evaluate fetal neurobehavioral development by evaluating fetal movements and facial expressions. Fetal facial movements and expressions such as blinking, a face without any expression, mouthing, scowling, smiling, sucking, tongue expulsion, and yawning can be evaluated by 4D ultrasound from the beginning of the 2nd trimester of pregnancy.2,14 Eye blinking (blinking) is a reflex response possibly related to brain function maturation and development that occurs with advancing gestation.14–18 Mouthing is the most frequent expression and is recognized as fetal brain maturation if it occurs together with non-rapid eye movement after 35 weeks of gestation.19 The frequency of scowling that might indicate suffering of the fetus in utero pain or stress20 increases with advancing gestation.21 Smiling might indicate a state of brain development performing complex facial movements.22,23 The correlation of an expressionless face and tongue expulsion with brain function is unclear.14 Yawning may be utilized as an index of fetal development.24,25 Therefore, it is important to investigate fetal facial expressions. There have been, however, no standard objective methods to evaluate fetal facial expressions.

Recently, artificial intelligence (AI) has advanced into the field of medicine. In different fields of obstetrics and gynecology, research works relevant to AI have been published.26–35 A well-trained AI classifier that can evaluate and classify fetal facial expressions would help investigate the development of the fetal central nervous system. The AI recognition of adult facial expressions has been investigated. Kim et al. reported the accuracy of the AI facial expression recognition was 0.965.36 Adult facial expressions can state human mental state and behavior and their analysis is available for marketing, healthcare, safety, environment, and social media.37

In this review article, we introduce the updated status of AI recognition of fetal facial expressions as a significant parameter for fetal brain function and suggest recommendations for future research on fetal brain development and function.

RECOGNITION OF FETAL FACIAL EXPRESSIONS USING AI

All data per fetus are divided into test/training/validation datasets at random in a ratio that is not fixed but commonly set to 0.20/0.64/0.16. In this way, training datasets, validation datasets, and non-overlapping test datasets are created.

The AI classifier is then designed. The AI classifier composed of convolutional neural network (CNN)38–43 for classifying categories is often used for image recognition. The CNN usually comprises...
layers with a combination of convolutional layers such as pooling layers, \(44\)-\(46\) linear layers, \(47\)-\(48\) flattened layers, \(49\) batch normalization layer, \(50\) rectified linear unit layers, \(51\),\(52\) and softmax layer \(53\),\(54\) that presents the probabilities of each category named confidence scores. Then, the category with the highest confidence score is determined as the AI classification category of each image. The AI classifier is trained using the training dataset with simultaneous validation by using the validation dataset. Before the AI training, the training and validation datasets are augmented by methods such as rotating the images. Data augmentation is often used, because image processing such as rotation, can result in different vector data in the same category.\(^{31}\)

The feasibility of the AI classifier is evaluated using the test dataset. Then, statistical values of the test dataset are then obtained, such as sensitivity, specificity, accuracy, receiver operating characteristic (ROC) curve, etc.

**Previous Studies on AI Recognition of Fetal Facial Expressions**

The development of the AI classifier with the original neural network architecture that recognizes and classifies images of fetal faces captured by sonography was reported by Miyagi et al.\(^{55}\) This pilot study was the first report on the recognition of fetal facial expressions by AI, to the best of our knowledge. The CNN architecture consisted of 13 layers; 2 convolution layers, 3 rectified linear unit layers, 2 pooling layers, 1 flatten layer, 3 linear layers, 1 batch normalization layer, and 1 softmax layer. The classifier could classify only five categories: blinking, face without any expression (neutral face), mouthing, scowling, and yawning, due to sample limitations for each category. The number of fetus/images were 93/922 and the number of test/validation/training datasets for creating the AI was 222/1,648/7,168. The accuracy for the test dataset was 0.985. The values of accuracy/sensitivity/specificity were 0.996/0.993/1.000, 0.992/0.986/1.000, 0.985/1.000/0.979, 0.996/0.888/1.000, and 1.000/1.000/1.000 for blinking, mouthing, neutral face, scowling, and yawning, respectively. Though the confidence score of the blinking category in the rated categories was 0.51 \(\pm\) 0.35 (mean \(\pm\) SD), the maximum values of the average of the confidence scores of other categories were over 0.85.

**Further Achievements in AI Recognition of Fetal Facial Expressions**

We introduce the improved AI classifier composed of the same neural network architecture by collecting more data for recognizing seven categories in this review. The number of fetus/images were 237/1,457 and the number of test/validation/training dataset for creating the AI was 251/1,536/11,248. The accuracy, the confidence scores, and the ROC curve of the AI fetal facial expression analysis were 0.996 as shown Figures 1 to 3, respectively. The accuracy/sensitivity/specificity values were 0.996/0.964/1.000, 1.000/1.000/1.000, 0.996/1.000/0.994, 1.000/1.000/1.000, 1.000/1.000/1.000, 1.000/1.000/1.000, and 1.000/1.000/1.000 for blinking, mouthing, neutral face, scowling, smiling, tongue expulsion, and yawning, respectively (Table 1). Other statistical values such as negative predictive value, positive predictive value, informedness, the area under the ROC curve, F1 score, markedness, and Matthews correlation coefficient were over 0.96 in all categories. Sample images classified by AI are shown in Figure 4.

**Limitations**

The following limitations of AI fetal facial expression recognition need to be considered. First, the AI cannot properly classify unknown images. Though seven facial categories were used, there would be other fetal facial expressions that are significant in investigating the development of the fetal central nervous system. The perfect classification of fetal facial expressions has not yet been established, possibly due to the long time needed to observe the fetal face and lack of consensus image classification by examiners. Such undefined and undiscovered images and categories would be needed to train AI for clinical practice and research in the future. Second, the feasibility of fetal facial expression recognition by AI depends on the supervised data by experienced examiners. Moreover, the anthropometric differences affected in fetal facial expression could strongly affect AI creation, this AI would not be feasibly used for different anthropometric fetuses directly. We believe, however, that similar algorithms would be available for other anthropometric fetuses. Third and last, although the AI showed quite good accuracy, there are still defective datasets such as for sucking that was rarely seen during the examination. The incidence of sucking was approximately 1\% in all cases.\(^{55}\)

More datasets are required, as in general, AI deep learning for the neural network is better with more datasets. The recognition frequencies and accuracies of each fetal facial expression related to gestational weeks by AI should also be analyzed.

**Further Perspective**

The advantage of multi-modalities for AI has been presented in the classification of the uterine cervical squamous epithelial lesion from colposcopy images combined with HPV types\(^{27}\) and the predicting live birth from blastocyst images combined with the conventional clinical embryo evaluation parameters.\(^{28}\) Therefore, fetal facial expressions can be classified by image and by incorporating images with gestational age and other associated parameters.
Established AI has no intrinsic bias for classifying images. Thus, AI can show objective findings regarding fetal facial expression recognition, which could advance research on the fetal central nervous system and brain development. The establishment of AI classification of fetal facial expressions could enable objective fetal neurodevelopment investigation by applying tests such as mini KANET that is used for predicting postnatal developmental disabilities. Further establishment of an advanced AI recognition for fetal facial expression using incorporated images with associated parameters will be able to reveal correlations between facial expressions and parameters such as fetal physical development, multiple pregnancies, parity, siblings, maternal personality, maternal disease, mental and physical development after birth, personality formation, score in school, intelligence, etc. Subsequent observation data of medical and social factors obtained in cohort studies or retrospective studies may aid mothers in the next generation in providing optimal treatment for their fetuses.

**Conclusion**

As the fetal facial expression is considered to be important for non-invasively investigating the fetal central nervous system and brain development, AI may be useful in this aspect with the advent of 4D ultrasound.

Fig. 2: The profiles of the confidence scores for each category per image (mean ± standard deviation) for the test dataset
Recognition of Fetal Facial Expressions by AI

Table 1: Statistic results of AI for classifying fetal facial expressions

| Statistic results                  | Eye blinking | Neutral face | Mouthing | Scowling | Smiling | Tongue expulsion | Yawning |
|-----------------------------------|-------------|-------------|----------|----------|----------|------------------|---------|
| True-positive number              | 27          | 88          | 90       | 9        | 4        | 7                | 25      |
| True-negative number              | 223         | 162         | 161      | 242      | 247      | 244              | 226     |
| False-positive number             | 0           | 1           | 0        | 0        | 0        | 0                | 0       |
| False-negative number             | 1           | 0           | 0        | 0        | 0        | 0                | 0       |
| Accuracy                          | 0.996       | 0.996       | 1.000    | 1.000    | 1.000    | 1.000            | 1.000   |
| Area under ROC curve              | 1.000       | 1.000       | 1.000    | 1.000    | 1.000    | 1.000            | 1.000   |
| F1 score                          | 0.982       | 0.994       | 1.000    | 1.000    | 1.000    | 1.000            | 1.000   |
| False discovery rate              | 0.000       | 0.011       | 0.000    | 0.000    | 0.000    | 0.000            | 0.000   |
| False-negative rate               | 0.036       | 0.000       | 0.000    | 0.000    | 0.000    | 0.000            | 0.000   |
| False-positive rate               | 0.000       | 0.006       | 0.000    | 0.000    | 0.000    | 0.000            | 0.000   |
| Informedness                      | 0.964       | 0.994       | 1.000    | 1.000    | 1.000    | 1.000            | 1.000   |
| Markedness                        | 0.996       | 0.989       | 1.000    | 1.000    | 1.000    | 1.000            | 1.000   |
| Matthews correlation coefficient  | 0.980       | 0.991       | 1.000    | 1.000    | 1.000    | 1.000            | 1.000   |
| Negative predictive value         | 0.996       | 1.000       | 1.000    | 1.000    | 1.000    | 1.000            | 1.000   |
| Positive predictive value (precision)  | 1.000 | 0.989       | 1.000    | 1.000    | 1.000    | 1.000            | 1.000   |
| Sensitivity (recall)              | 0.964       | 1.000       | 1.000    | 1.000    | 1.000    | 1.000            | 1.000   |
| Specificity                       | 1.000       | 0.994       | 1.000    | 1.000    | 1.000    | 1.000            | 1.000   |

ROD, receiver operating characteristic

Fig. 3: The receiver operating characteristic (ROC) curve of AI for the test dataset. The area under the ROC curve was 1.0

References

1. Hata T, Dai SY, Marumo G. Ultrasound for evaluation of fetal neurobehavioural development: from 2-D to 4-D ultrasound. Inf Child Dev 2010;19:99–118. DOI: 10.1002/icd.659.
2. Hata T, Kanenishi K, Hanaoka U, et al. HDI live and 4D ultrasound in the assessment of fetal facial expressions. Donald School J Ultrasound Obstet Gynecol 2015;9:44–50. DOI: 10.5005/jp-journals-10009-1388.
3. Hata T. Current status of fetal neurodevelopmental assessment: 4D ultrasound study. J Obstet Gynecol Res 2016;42:1211–1221. DOI: 10.1111/jog.13099.
4. Nijhuis JG. Fetal behavior. Neurobiol Aging 2003;24(Suppl. 1):S41–S46. DOI: 10.1016/S0197-4580(03)00054-X.
5. Prechtl HF. State of the art of a new functional assessment of the young nervous system: an early predictor of cerebral palsy. Early Hum Dev 1997;50:1–11. DOI: 10.1016/S0378-3782(97)00088-1.
6. de Vries JIP, Visser GHA, Prechtl HFR. The emergence of fetal behaviour. I. Qualitative aspects. Early Hum Dev 1982;7:301–322. DOI: 10.1016/0378-3782(82)90033-0.
7. de Vries JIP, Visser GHA, Prechtl HFR. The emergence of fetal behaviour. II. Quantitative aspects. Early Hum Dev 1985;12:99–120. DOI: 10.1016/0378-3782(85)90174-4.
8. Prechtl HF. Qualitative changes of spontaneous movements in fetus and preterm infant are a marker of neurological dysfunction. Early Hum Dev 1990;23:151–158. DOI: 10.1016/0378-3782(90)90017-7.
9. Prechtl HF, Einspieler C. Is neurological assessment of the fetus possible? Eur J Obstet Gynecol Reprod Biol 1997;75:81–84. DOI: 10.1016/S0301-2115(97)00197-8.
10. Kuno A, Akiyama M, Yamashiro C, et al. Three-dimensional sonographic assessment of fetal behavior in the early second trimester of pregnancy. J Ultrasound Med 2001;20:1271–1275. DOI: 10.1046/j.1469-0705.2001.abs20-7.x.
11. Hata T. Fetal face as predictor of fetal brain. Donald School J Ultrasound Obstet Gynecol 2018;12(1):56–59.
12. Kurjak A, Miskovic B, Stanojevic M, et al. New scoring system for fetal neurobehavior assessed by three- and four-dimensional sonography. J Perinat Med 2008;36:73–81. DOI: 10.1515/JPM.2008.007.
13. Stanojevic M, Talic A, Miskovic B, et al. An attempt to standardize Kurjak’s antenatal neurodevelopmental test: Osaka consensus statement. Donald School J Ultrasound Obstet Gynecol 2011;5:317–329. DOI: 10.5005/jp-journals-10009-1209.
14. AboEllail MAM, Hata T. Fetal face as important indicator of fetal brain function. J Perinat Med 2017;45:729–736. DOI: 10.1515/jpm-2016-0377.
15. Bodfish J, Powell S, Golden R, et al. Blink rate as an index of dopamine function in adults with mental retardation and repetitive movement disorders. Am J Ment Retard 1995;99:335–344.
16. Kleven MS, Koek W. Differential effects of direct and indirect dopamine agonists on eye blink rate in cynomolgus monkeys. J Pharmacol Exp Ther 1996;279:1211–1219.
17. Driebach G, Muller J, Goschke T, et al. Dopamine and cognitive control: the influence of spontaneous eyeblink rate and dopamine gene polymorphism on perseveration and distractibility. Behav Neurosci 2005;119:483–490.
18. Colzato LS, van den Wildenberg WPM, van Wouwe NC, et al. Dopamine and inhibitory action control: evidence from spontaneous eye blink rates. Exp Brain Res 2009;196:467–474.
Recognition of Fetal Facial Expressions by AI

Donald School Journal of Ultrasound in Obstetrics and Gynecology, Volume 15 Issue 3 (July–September 2021)

19. Horimoto N, Koyanagi T, Nagata S, et al. Concurrence of mouthing movement and rapid eye movement/non-rapid eye movement phases with advance in gestation of the human fetus. Am J Obstet Gynecol 1989;161:344–351.
20. Hata T, Kanenishi K, AboEllail MAM, et al. Fetal consciousness 4D ultrasonography study. Donald School J Ultrasound Obstet Gynecol 2015;9:471–474. DOI: 10.5005/jp-journals-10009-1434.
21. Reissland N, Francis B, Mason J. Can healthy fetuses show facial expressions of “pain” or “distress”? PLOS One 2013;8:e65530.
22. Kawakami F, Yanaihara T. Smiles in the fetal period. Infant Behav Dev 2012;35:466–471.
23. Reissland N, Francis B, Mason J, et al. Development of fetal yawn compared with non-yawn mouth openings from 24–36 weeks gestation. PLoS One 2012;7:e50569.
24. Walusinski O, Kurjak A, Andonotopo W, et al. Fetal yawning assessed by 3D and 4D sonography. Ultrasound Rev Obstet Gynecol 2005;5:210–217.
25. Reissland N, Francis B, Manson J. Development of fetal yawn compared with non-yawn mouth openings from 24–36 weeks gestation. PLoS One 2012;7:e50569.
26. Miyagi Y, Miyake T. Potential of artificial intelligence for estimating Japanese fetal weights. Acta Medica Okayama 2020;74:483–493. DOI: 10.18926/AMO/61207.
27. Miyagi Y, Takehara K, Nagayasu Y, et al. Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images combined with HPV types. Oncol Lett 2020;19:1602–1610. DOI: 10.3892/ol.2019.11214.
28. Miyagi Y, Takehara K, Miyake T. Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images. Mol Clin Oncol 2019;11:583–589. DOI: 10.3892/mco.2019.1932.
29. Miyagi Y, Habara T, Hirata R, et al. Predicting a live birth by artificial intelligence incorporating both the blastocyst image and conventional embryo evaluation parameters. Artif Intell Med Imaging 2020;1:94–107. DOI: 10.35711/aimi.v1.i3.94.
30. Miyagi Y, Habara T, Hirata R, et al. Feasibility of artificial intelligence for predicting live birth without aneuploidy from a blastocyst image. Reprod Med Biol 2019;18:204–211. DOI: 10.1002/mbb.12267.
31. Miyagi Y, Habara T, Hirata R, et al. Feasibility of deep learning for predicting live birth from a blastocyst image in patients classified by age. Reprod Med Biol 2019;18:190–203. DOI: 10.1002/mbb.12266.
32. Miyagi Y, Habara T, Hirata R, et al. Feasibility of predicting live birth by combining conventional embryo evaluation with artificial intelligence applied to a blastocyst image in patients classified by age. Reprod Med Biol 2019;18:344–356. DOI: 10.1002/mbb.12284.
33. Miyagi Y, Fujiwara K, Oda T, et al. Studies on development of new method for the prediction of clinical trial results using compressive sensing of artificial intelligence. In: Theory and Practice of Mathematics and Computer Science Ferreira MAM, ed., Hooghly, West Bengal, India: Book Publisher International; 2020, pp. 101–108. DOI: 10.9734/bpi/tpmcs/v2.
34. Miyagi Y, Fujiwara K, Oda T, et al. Development of new method for the prediction of clinical trial results using compressive sensing of artificial intelligence. J Biostat Biometric App 2018;3:202.
35. Miyagi Y, Tada K, Yasuhi i, et al. New method for determining fibrinogen and FDP threshold criteria by artificial intelligence in cases of massive hemorrhage during delivery. J Obstet Gynaecol Res 2020;46:256–265. DOI: 10.1111/jog.14166.

Fig. 4: Sample images classified by AI. The accuracy for classifying fetal facial expression for seven categories was 0.996, which was almost equivalent to human classification.
Recognition of Fetal Facial Expressions by AI

36. Kim J, Kim B, Roy PP, et al. Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. IEEE Access 2019;7:41273–41285. DOI: 10.1109/ACCESS.2019.2907327.

37. Dixit AN, Kasbe T. A Survey on Facial Expression Recognition using Machine Learning Techniques. In 2nd International Conference on Data, Engineering and Applications (IDEA). 2020, 1-6. 10.1109/IDEA49133.2020.9170706.

38. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35:1798–1828. DOI: 10.1109/TPAMI.2013.50.

39. LeCun YA, Bottou L, Orr GB, et al. Efficient backprop. In: Neural networks: tricks of the trade Montavon G, Orr GB, Müller KR, ed., Heidelberg, Berlin: Springer; 2012. pp. 9–48. DOI: 10.1007/978-3-642-35289-8_3.

40. LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proc IEEE 1998;86:2278–2324. DOI: 10.1109/5.726791.

41. LeCun Y, Boser B, Denker JS, et al. Back propagation applied to handwritten zip code recognition. Neural Comput 1989;1:541–551. DOI: 10.1162/neco.1989.1.4.541.

42. Serre T, Wolf L, Bileschi S, et al. Robust object recognition with cortex-like mechanisms. IEEE Trans Pattern Anal Mach Intell 2007;29:411–426. DOI: 10.1109/TPAMI.2007.36.

43. Wiatowski T, Bölcskei H. A mathematical theory of deep convolutional neural networks for feature extraction. IEEE Trans Inf Theory 2017;64:1845–1866. DOI: 10.1109/TIT.2017.2776228.

44. Ciresan DC, Meier U, Masci J, et al.. Flexible, High Performance Convolutional Neural Networks for Image Classification. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 2011: 1237-1242.

45. Scherer D, Müller A, Behnke S. Evaluation of pooling operations in convolutional architectures for object recognition. In: Artificial neural networks – ICANN 2010. Lecture notes in computer science Diamantaras K, Duch W, Iliadis LS, ed., Berlin, Heidelberg: Springer; 2010. 92–101. DOI: 10.1007/978-3-642-15825-4_10.

46. Huang FJ, LeCun Y. Large-Scale Learning with Svm and Convolutional for Generic Object Categorization. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, USA. IEEE, 2006: 284–291 10.1109/CVPR.2006.164.

50. Ioff S, Szegedy C. Batch Normalization: Accelerating Deep Neural Training by Reducing Internal Covariate Shift. https://arxiv.org/abs/1502.03167v3.

51. Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Neural Networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS) 2011. AISTATS; Lauderdale, USA, 2011; 315-323.

52. Nair V, Hinton GE, Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th international conference on machine learning (ICML-10). Omnipress; Haifa, Israel, 2010: 807-814.

53. Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems. 2012: 1097-1105. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

54. Bridle JS. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: Neurocomputing Soulé F, Hérault J, ed., Berlin, Heidelberg: Springer; 1990. pp. 227–236. DOI: 10.1007/978-3-642-76153-9_28.

55. Miyagi Y, Hata T, Bouno S, et al. Recognition of facial expression of fetuses by artificial intelligence (AI). J Perinat Med 2021;49(5):596–603. DOI: 10.1515/jpm-2020-0537.

56. Hata T, Kanenishi K, Mori N, et al. Mini KANET: simple fetal antenatal neurodevelopmental test. Donald School J Ultrasound Obstet Gynecol 2019;13(2):59–63.