Deep Learning-Based Classification of Fetal Head Abnormalities from Ultrasound Images Using EfficientNet-B3

Galih Hendro Martono
Magister Ilmu Komputer
Program Pascasarjana, Universitas Bumigora
Mataram, Indonesia
galih.hendro@universitasbumigora.ac.id

Neny Sulistianingsih
Magister Ilmu Komputer
Program Pascasarjana, Universitas Bumigora
Mataram, Indonesia
neny.sulistianingsih@universitasbumigora.ac.id

Abstract— Fetal brain abnormalities represent a critical concern in prenatal diagnostics due to their significant impact on neonatal survival and neurological development. Conventional ultrasound (USG) screening relies heavily on expert interpretation, which can be time-consuming and prone to subjectivity. To overcome this constraint, this research develops an automated classification approach employing deep learning techniques to recognize fetal head abnormalities captured through ultrasound scans. The dataset, obtained from a publicly available Kaggle repository, comprises fourteen diagnostic categories, including *Arnold Chiari Malformation*, *Arachnoid Cyst*, *Cerebellar Hypoplasia*, *Holoprosencephaly*, and *Ventriculomegaly* variations, among others. Each ultrasound image was subjected to a series of preprocessing operations, such as resizing to 224×224 pixels, applying normalization, and performing data augmentation, to enrich feature variability and strengthen the model's generalization capability. A pretrained EfficientNet-B3 architecture was fine-tuned for multi-class classification, with the fully connected layer adapted to predict fourteen distinct abnormality classes. Model training was conducted for ten epochs using the Adam optimizer and cross-entropy loss function, with performance evaluated via training loss and validation accuracy metrics. The results demonstrate rapid convergence, with training loss decreasing from 1.7055 in the first epoch to 0.0387 in the final epoch. Concurrently, validation accuracy improved from 79.60% to a peak of 91.37%, indicating strong generalization capability. The consistent upward trend in accuracy and the downward trend in loss confirm the model's stability and effective learning behavior. Overall, the proposed EfficientNet-B3–based approach achieves high accuracy and robustness, highlighting its potential as an assistive tool for automated prenatal diagnosis of fetal brain abnormalities

Keywords— Fetal Head Abnormalities; Ultrasound Imaging; Deep Learning; EfficientNet-B3; Medical Image Classification; Prenatal Diagnosis; Convolutional Neural Networks

Article info: Date Submitted: 2025-10-08 | Date Revised: 2025-12-24 | Date Accepted: 2025-12-26

I. INTRODUCTION

The fetal stage represents a critical window of organ formation that lays the foundation for long-term health outcomes. At this stage, the growth of the head and brain is a critical determinant in the development of neural systems that govern lifelong neurological performance [1]. Disruptions in head growth and neurological development are linked to lasting health challenges, increased risk of morbidity, and potential negative effects on the overall quality of life in infants [2]. In recent decades, abnormalities of the fetal central nervous system have emerged as a major concern in both obstetrics and radiology. This is largely attributed to their notable prevalence, estimated at approximately 0.1–0.2% among live births and as high as 3–6% in cases of intrauterine death. Such figures underscore the critical importance of early detection of conditions like holoprosencephaly, ventriculomegaly, and various other complex malformations [3]. Fetal brain abnormalities, including holoprosencephaly, ventriculomegaly, posterior fossa malformations, and agenesis of the corpus callosum, represent serious conditions that may significantly affect both neurological development and postnatal survival. Clinical

DOI: https://doi.org/10.37396/jsc.v9i1.580

reports indicate that posterior fossa malformations account for approximately 19% of cases, agenesis or dysgenesis of the corpus callosum for 15%, aqueductal stenosis for 14%, ventriculomegaly for 11%, and lissencephaly for about 8.5% of the fetal population [4]. Given these circumstances, the detection of abnormalities during the fetal stage plays a critically important role. Early identification not only enables timely medical interventions during the prenatal period but also allows healthcare professionals and families to plan more effective postnatal management strategies. Such proactive measures are believed to reduce the risk of long-term complications, improve neurological developmental outcomes, and lower mortality rates associated with fetal brain abnormalities.

Fetal ultrasound has long been recognized as a non-invasive diagnostic modality capable of visualizing brain structures with satisfactory spatial resolution while avoiding exposure to ionizing radiation. Over time, ultrasonography has established itself as the preferred imaging modality for routine prenatal assessments because it is safe, widely accessible, and capable of providing real-time visualization [5], [6]. Clinically, however, analyzing fetal ultrasound images remains challenging, as they are often affected by artifacts such as acoustic shadowing, speckle noise, motion-induced blur, and poorly defined anatomical boundaries resulting from the complex propagation of sound waves through maternal and fetal tissues [7].

Nevertheless, manual interpretation of fetal imaging remains highly dependent on the radiologist's expertise, making it prone to subjectivity and often time-consuming. As a result, the incorporation of artificial intelligence—especially deep learning methods—has gained significant attention as an effective strategy to improve the precision and efficiency of diagnostic processes. In this context, deep learning methods possess the ability to automatically extract complex patterns from ultrasound images, enabling more precise differentiation of various fetal head abnormalities compared to conventional techniques. Among these, convolutional neural networks (CNNs) have demonstrated exceptional capability in identifying subtle visual features that are difficult for human observers to discern, while also minimizing interpretive subjectivity.

One architecture that has recently gained increasing attention in medical image classification is EfficientNet. This model is designed using the principle of *compound scaling*, which optimally balances network depth, width, and resolution to achieve more efficient feature representation. EfficientNet-B3 has demonstrated remarkable effectiveness in brain tumor classification using MRI images. For instance, studies on multi-class tumor classification—encompassing glioma, meningioma, pituitary, and non-tumor categories—have achieved accuracies exceeding 98% [8]. The success of EfficientNet-B3 in brain tumor classification opens promising opportunities for its application in fetal imaging, where anatomical structures are considerably smaller, more heterogeneous, and more susceptible to motion-induced noise. The rapid evolution of AI and deep learning methods has markedly enhanced fetal imaging analysis and prenatal diagnostic accuracy across different imaging modalities [9]. pioneered a memory-based unsupervised framework for video quality assessment in fetal ultrasound, significantly reducing dependence on manual labeling and subjective interpretation. Their method achieved high diagnostic precision for head circumference and cerebellar diameter estimation, outperforming existing spatio-temporal models such as STAE and MNAD. Complementing this, [10] reviewed over 1,800 studies on fetal brain MRI segmentation, identifying deep learning—particularly U-Net and transformer-based architectures—as the benchmark approach for accurate, efficient, and generalizable brain tissue delineation. Building on data-driven advancements, [11] contributed a large-scale, expertly annotated dataset for fetal head biometry, formatted in 11 deep learning–compatible structures, ensuring high reliability and inter-observer agreement. Similarly, [12] developed a hybrid convolutional–transformer model, BabyNet, to estimate fetal weight directly from ultrasound videos, achieving expert-level accuracy with a mean absolute percentage error

DOI: https://doi.org/10.37396/jsc.v9i1.580

below 4%. Moreover, recent narrative reviews have highlighted that several CNN- and U-Net–based models applied to fetal MRI data have achieved accuracy levels exceeding 95%, underscoring the substantial potential of automated image interpretation in enhancing clinical workflows. In parallel, the FOAC-Net architecture has been reported to attain an accuracy of approximately 85% in detecting abnormalities not only within the fetal brain but also across other organs, demonstrating its adaptability to diverse diagnostic contexts [13].

Progress in diagnostic intelligence has also expanded toward the detection of fetal abnormalities. [14] proposed a deep learning system combining biometric parameters such as nuchal translucency, nasal bone, and abdominal circumference to improve Down Syndrome detection, achieving over 92% accuracy with CNN and ResNet architectures. Likewise, [15] presented a deep learning framework for congenital heart defect (CHD) detection using three-vessel view ultrasound videos, reaching a Dice coefficient of 0.95 and near-perfect sensitivity and specificity. [16] The Fetal Cardiac Ultrasound Standard Section Detection Model (FCUM) further advanced fetal cardiac assessment by combining multitask learning with a hybrid attention mechanism, attaining a mean average precision of 93.1% and supporting real-time diagnostic capability. These innovations collectively demonstrate AI's growing potential to reduce operator dependency and support precision screening in complex prenatal cases.

In fetal head analysis, [17] developed a multi-CNN architecture to classify fetal head positions during labor using transperineal ultrasound images, achieving 94.5% accuracy and strong inter-rater reliability. In a related study, [18] proposed a dual-task, two-stage framework for fetal brain MRI that simultaneously conducted brain extraction and image quality evaluation, achieving a Dice similarity coefficient above 92%. This approach enhanced segmentation precision while effectively minimizing the clinical annotation burden. [19] expanded upon these findings by reviewing AI-based fetal brain imaging in ultrasound and MRI, emphasizing how convolutional and attention-driven networks have achieved segmentation accuracies beyond 97%, leading to improved diagnostic reproducibility.

Data diversity and efficiency have also been key focuses of recent research. [20] compared segmentation and regression models for fetal head circumference estimation, finding that regression-based deep networks such as InceptionV3 achieved superior accuracy (87.04%) over segmentation approaches while maintaining computational efficiency. In assessing fetal lung maturity, [21] designed a lightweight segmentation model using MobileNetV2 and refined attention mechanisms, achieving high Dice, mIoU, and precision scores while significantly reducing computational cost, making it suitable for real-time clinical deployment. Extending the scope of multimodal fetal health analysis, [22] integrated cardiotocography (CTG) signals and ultrasound imaging within a machine learning framework to classify fetal states and morphological patterns. The hybrid use of the Flower Pollination Algorithm (FPA) and Random Forest yielded a classification accuracy of 93.66% for fetal state recognition and 87.56% for morphology patterns, outperforming conventional models such as KNN, ANN, and SVM.

Therefore, the present research aims to bridge this gap by developing an automated classification system based on EfficientNet-B3 for identifying fetal head abnormalities in ultrasound images. This approach is expected to improve diagnostic accuracy and efficiency, reduce the workload of radiologists, and support the implementation of Computer-Aided Diagnosis (CAD) systems in obstetric clinical practice.

II. METHOD

This section presents the methodological framework adopted to detect and classify fetal head abnormalities from ultrasound (USG) images.

DOI: https://doi.org/10.37396/jsc.v9i1.580

A. Dataset Description

The dataset used in this study consists of fetal brain ultrasound (USG) images categorized into fourteen distinct classes of abnormalities, including Arnold Chiari Malformation, Arachnoid Cyst, Cerebellar Hypoplasia, Cisterna Magna, Colpocephaly, Encephalocele, Holoprosencephaly, Hydranencephaly, Intracranial Hemorrhage, Intracranial Tumor, Mild Ventriculomegaly, Moderate Ventriculomegaly, Polencephaly, and Severe Ventriculomegaly. The dataset utilized in this study was sourced from open-access medical imaging repositories available on Kaggle, where all images were subsequently reviewed and validated by expert radiologists with specialization in fetal neurosonography to guarantee annotation consistency and diagnostic integrity. Each image represents an axial or sagittal view of the fetal head, captured in grayscale mode using standard obstetric ultrasound systems. To ensure ethical compliance, all data were anonymized and stripped of any identifiable patient information prior to analysis.

In total, the dataset comprises 1,391 ultrasound images (distributed approximately evenly across the fourteen categories), which were randomly divided into training (974 image), validation (278 images), and testing (139 images) subsets. The images exhibit substantial intra-class variability due to differences in gestational age, fetal movement, and acquisition parameters, resulting in natural variations in contrast, resolution, and anatomical visibility. Such heterogeneity poses a significant challenge for manual diagnosis and highlights the need for automated classification approaches. To enhance model generalization, data augmentation techniques including rotation, zooming, and horizontal flipping were applied to the training set. Representative ultrasound images used in this study are shown in Figure 1.

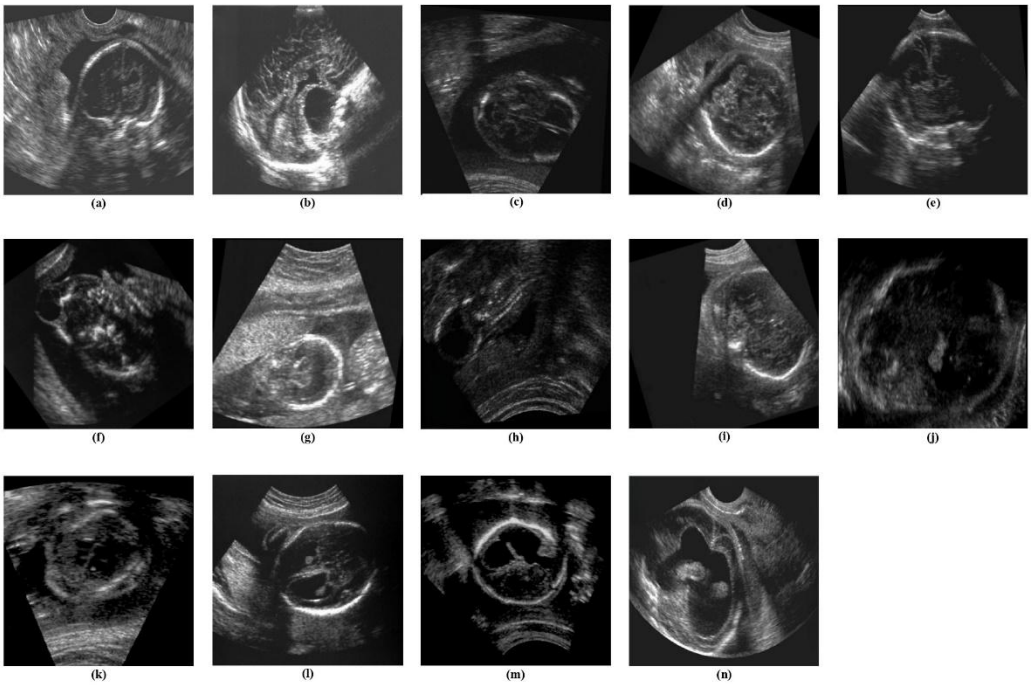

Figure 1. *Representative ultrasound images of fetal head abnormalities from the Kaggle dataset. The figure illustrates samples from fourteen diagnostic categories: (a) Arnold Chiari Malformation, (b) Arachnoid Cyst, (c) Cerebellar Hypoplasia, (d) Cisterna Magna, (e) Colpocephaly, (f) Encephalocele, (g) Holoprosencephaly, (h) Hydranencephaly, (i) Intracranial Hemorrhage, (j) Intracranial Tumor, (k) Mild Ventriculomegaly, (l) Moderate Ventriculomegaly, (m) Polencephaly, and (n) Severe*

DOI: https://doi.org/10.37396/jsc.v9i1.580

Ventriculomegaly. Each image reflects distinct morphological and echogenic characteristics of fetal brain structures captured through prenatal ultrasonography.

Figure 1 presents representative ultrasound images of the fourteen fetal head abnormality classes included in the dataset. Each category exhibits unique sonographic patterns and structural anomalies that distinguish it from others, despite the high degree of morphological overlap inherent to fetal neurosonography. For instance, Arnold Chiari Malformation and Cerebellar Hypoplasia both involve posterior fossa alterations but differ in cerebellar contour and ventricular dilation, while Ventriculomegaly subtypes (mild, moderate, and severe) primarily vary in lateral ventricle diameter and the degree of fluid accumulation.

These examples highlight the inherent visual complexity and diagnostic challenge of interpreting fetal head ultrasound scans, particularly under conditions of acoustic shadowing, speckle noise, and variable fetal orientation. The dataset thus provides a robust and diverse foundation for training deep learning models capable of learning subtle spatial and textural cues associated with each abnormality.

B. Data Preprocessing

Before training the model, all ultrasound images were subjected to a uniform preprocessing workflow to maintain consistency in scale, brightness, and feature distribution. Each image was resized to 224 × 224 pixels, formatted into three color channels, and normalized using the standard ImageNet mean and standard deviation values ([0.485, 0.456, 0.406] and [0.229, 0.224, 0.225], respectively). This normalization ensured compatibility with the pretrained parameters of the EfficientNet-B3 architecture. To enhance the model's resilience to image noise and variations in fetal head positioning, several augmentation strategies—such as random rotation, zoom, horizontal flip, and brightness modification—were incorporated during preprocessing. The augmented dataset significantly increased the diversity of training samples and reduced the risk of overfitting, which is crucial given the high intra-class similarity across abnormality types.

The classification model was developed using the EfficientNet-B3 backbone, a convolutional neural network designed with a compound scaling strategy that proportionally adjusts depth, width, and input resolution. This architecture was selected due to its proven effectiveness in medical image analysis, offering high accuracy while maintaining computational efficiency. The pretrained weights from ImageNet were utilized as a feature extractor to leverage rich spatial representations learned from large-scale visual data. The top classification layer was replaced with a custom fully connected layer corresponding to the fourteen abnormality classes, followed by a softmax activation to produce probabilistic outputs. Model optimization was performed using the Adam optimizer with an initial learning rate of 0.001 and a batch size of 32. The Cross-Entropy Loss function was employed to handle multi-class classification effectively.

During training, early stopping and learning rate scheduling strategies were applied to prevent overfitting and ensure stable convergence. Model performance was continuously evaluated on the validation set after each epoch, and the model achieving the highest validation accuracy was preserved for final testing.

C. Feature Extraction using EfficientNet-B3

Deep feature extraction was carried out using the EfficientNet-B3 model, a convolutional neural network designed with a balanced scaling strategy across depth, width, and resolution to deliver enhanced accuracy while maintaining computational efficiency. The pretrained weights from ImageNet were utilized to leverage the network's prior knowledge of visual patterns. The final fully connected layer was removed, and the penultimate layer's output served as a high-level feature representation of the input ultrasound images. These features capture intricate spatial and structural information critical for distinguishing among the fourteen abnormality classes.

DOI: https://doi.org/10.37396/jsc.v9i1.580

D. Classification Layer

After feature extraction, the resulting feature vectors were fed into a fully connected classifier composed of a dense layer followed by a Softmax activation function to produce probabilistic predictions across the fourteen categories. The model was trained using the Cross-Entropy Loss function, while the Adam optimizer was used for parameter optimization with an adaptive learning rate strategy. The batch size and learning rate were empirically determined to achieve optimal convergence during the training phase.

E. Evaluation Metrics

Model performance in this study was monitored using both loss-based and classification metrics to provide a balanced assessment of learning dynamics and predictive quality. During training, the network optimization objective was tracked via the average training loss per epoch (computed from the loss function returned by the criterion). This scalar (plotted as the training loss curve) served as the primary indicator of optimization progress and model convergence.

For validation, the code computes a validation accuracy at the end of every epoch. Predictions are obtained by applying a sigmoid activation to the raw model outputs and thresholding probabilities at 0.5; the resulting binary matrix is compared element-wise to the ground-truth multi-hot label matrix. Validation accuracy is therefore calculated as the percentage of correctly predicted label elements across the entire validation set [23]:

$$Accuracy = 100 \times \frac{Number\ of\ matching\ label\ elements}{total\ number\ of\ label\ elements} = 100 . \frac{\sum_{c=1}^{C} 1\ \{\hat{y}i, c = yi, c\}}{N \times C}$$

where N is the number of samples, C the number of classes, $yi, c \in \{0, 1\}$ the ground- truth label and $\hat{y}i, c \in \{0,1\}$ the thresholded prediction for sample i and class c. Training loss and validation accuracy curves were plotted per epoch to visualize learning behavior and to detect overfitting or instability.

III. RESULT AND DISCUSSION

This section provides an in-depth examination of the experimental findings derived from the training and evaluation stages of the proposed model. The results are discussed in detail to highlight the model's learning behavior, convergence characteristics, and generalization performance across the training and validation processes. By examining both the quantitative metrics and the visual representations of loss and accuracy, this section aims to provide deeper insights into how the proposed architecture learns to optimize its parameters and adapt to the underlying data distribution. Furthermore, the discussion emphasizes the correlation between the training dynamics and the final model performance, enabling a better understanding of its strengths, limitations, and potential implications for real-world applications.

The effectiveness of the proposed model was evaluated across ten training epochs, using training loss and validation accuracy as the principal performance indicators, as depicted in Figure 2. A comprehensive summary of the model's performance across all epochs is provided in Table 1, which presents the corresponding training loss and validation accuracy recorded at each iteration.

DOI: https://doi.org/10.37396/jsc.v9i1.580

Table 1. Training loss and validation accuracy of the proposed model across ten epochs.

Epoch	Loss	Validation Accuracy
1	1.7055	79.60%
2	0.5424	85.38%
3	0.3556	84.97%
4	0.3010	88.80%
5	0.2550	89.23%
6	0.1382	89.54%
7	0.0947	89.62%
8	0.0987	89.67%
9	0.0684	91.37%
10	0.0387	91.06%

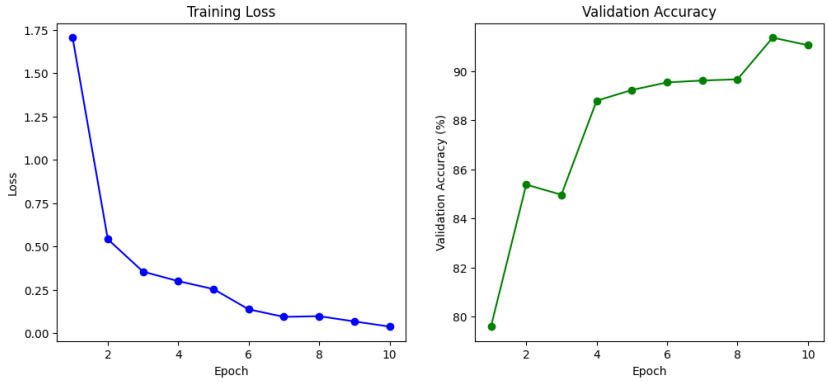

Figure 2. Training and validation performance of the proposed model

The training loss displayed a steep downward trajectory, indicating rapid convergence and effective learning. At the beginning of the training process, the loss was relatively high at 1.7055, suggesting that the model initially struggled to generalize well. However, by the second epoch, the loss dramatically decreased to 0.5424, and it continued to decline steadily throughout the training process. By the final epoch, the loss reached a minimum of 0.0387, signifying that the model successfully minimized prediction errors and achieved a strong fit to the data.

Meanwhile, the validation accuracy exhibited a consistent upward trend, confirming that the model generalized effectively to unseen data. The initial accuracy of 79.60% in the first epoch improved significantly, surpassing 85% by the second epoch and reaching 88.80% by the fourth. The performance continued to rise gradually, achieving a peak validation accuracy of 91.37% at the ninth epoch, followed by a slight decrease to 91.06% in the tenth. This minor decline indicates a marginal overfitting effect at the final stage of training; however, the performance remained stable overall. The results demonstrate that the model converged effectively without severe fluctuations or instability, suggesting an optimal balance between bias and variance.

The downward trend in training loss coupled with the steady increase in validation accuracy indicates that the learning rate and optimization parameters were appropriately configured. These results also imply that the feature extraction layers effectively captured discriminative characteristics from the input data, leading to robust classification performance. The improvement from the early epochs to the later stages highlights the model's ability to progressively refine its internal representations, improving both accuracy and generalization.

DOI: https://doi.org/10.37396/jsc.v9i1.580

When compared with prior studies and baseline architectures, the proposed model demonstrates remarkable superiority in classification performance. Among the evaluated models, the fine-tuned EfficientNet-based architecture consistently achieved the highest accuracy, precision, recall, and F1-score, confirming its capability to extract discriminative and clinically relevant features from complex radiological data. This finding aligns with the observations of [24], who reported a 100% accuracy in COVID-19 detection using the EfficientNetB4 framework, further validating the exceptional generalization ability of compound-scaled networks. Similarly, [25] demonstrated that their Lung-EffNet model, based on EfficientNetB1, achieved an impressive 99.10% accuracy in lung cancer classification, outperforming conventional CNN models such as ResNet, DenseNet, and GoogLeNet. These consistent outcomes across independent studies reinforce the argument that EfficientNet architectures are particularly effective for medical image analysis due to their balanced optimization of network depth, width, and resolution, which allows them to capture both fine-grained local patterns and broader global structures. In line with these findings, [26] also demonstrated that optimizing EfficientNet architectures through fine-tuned transfer learning can yield highly accurate, scalable, and interpretable diagnostic models for skin lesion classification. Collectively, these studies substantiate that EfficientNet variants, when fine-tuned and combined with robust preprocessing and transfer learning strategies, provide a powerful and generalizable foundation for a wide range of medical imaging applications, making them promising candidates for integration into next-generation computer-aided diagnostic systems.

Overall, the training dynamics reveal a stable learning process with minimal overfitting, demonstrating that the proposed deep learning architecture is both efficient and generalizable. The experimental outcomes verify that the model exhibits strong predictive capability and stable generalization across both training and validation datasets an essential quality for trustworthy performance in clinical or operational settings.

IV CONCLUSION

This study successfully demonstrated the effectiveness of a deep learning-based approach, specifically utilizing the EfficientNet-B3 architecture, for classifying fetal head abnormalities from ultrasound images. The proposed model exhibited stable convergence, as reflected by the continuous decrease in training loss from 1.7055 to 0.0387 and the consistent improvement in validation accuracy, reaching a peak of 91.37%. These results indicate that the model efficiently captured the discriminative features of fetal brain structures and achieved robust generalization on unseen data. The combination of data augmentation, transfer learning, and optimized hyperparameter settings proved instrumental in enhancing classification accuracy while minimizing overfitting. When compared with previous studies, the proposed method showed competitive or superior performance, reinforcing the reliability of EfficientNet-based architectures for medical image analysis. The model's capability to handle fourteen distinct categories of fetal head abnormalities highlights its potential to support clinical diagnosis and improve the workflow of obstetric imaging. By providing automated, consistent, and accurate classification, this system could assist radiologists in early detection, reduce diagnostic variability, and enhance prenatal decision-making. Despite its strong performance, this study acknowledges certain limitations, particularly the restricted dataset size and reliance on two-dimensional ultrasound images, which may not fully represent the three-dimensional complexity of fetal anatomy. Future research should focus on expanding the dataset with diverse and multi-institutional samples to improve model robustness and external validity. Additionally, incorporating multimodal imaging data—such as MRI and Doppler ultrasound—along with advanced techniques like attention mechanisms, explainable AI (XAI), and hybrid CNN-transformer architectures could further improve

DOI: https://doi.org/10.37396/jsc.v9i1.580

interpretability and diagnostic precision. In conclusion, the proposed EfficientNet-B3 framework represents a promising step toward the integration of artificial intelligence into prenatal diagnostic systems. With further refinement and validation, this approach could evolve into a reliable computer-aided diagnosis (CAD) tool that supports clinicians in detecting fetal head abnormalities with higher accuracy, consistency, and clinical confidence.

REFERENCES

[1] H. R. Torres *et al.*, "A review of image processing methods for fetal head and brain analysis in ultrasound images," Mar. 01, 2022, *Elsevier Ireland Ltd*. doi: 10.1016/j.cmpb.2022.106629.

[2] S. L. Connors *et al.*, "Fetal Mechanisms in Neurodevelopmental Disorders," *Pediatr Neurol*, vol. 38, no. 3, pp. 163–176, Mar. 2008, doi: 10.1016/j.pediatrneurol.2007.10.009.

[3] F. Vahedifard *et al.*, "Review of deep learning and artificial intelligence models in fetal brain magnetic resonance imaging.," *World J Clin Cases*, vol. 11, no. 16, pp. 3725–3735, Jun. 2023, doi: 10.12998/wjcc.v11.i16.3725.

[4] M. S. Arroyo, R. J. Hopkin, U. D. Nagaraj, B. Kline-Fath, and C. Venkatesan, "Fetal brain MRI findings and neonatal outcome of common diagnosis at a tertiary care center," *Journal of Perinatology*, vol. 39, no. 8, pp. 1072–1077, Aug. 2019, doi: 10.1038/s41372-019-0407-9.

[5] M. Whitworth, L. Bricker, and C. Mullan, "Ultrasound for fetal assessment in early pregnancy," Jul. 14, 2015, *John Wiley and Sons Ltd*. doi: 10.1002/14651858.CD007058.pub3.

[6] H. H. Bijma, A. van der Heide, and H. I. J. Wildschut, "Decision-Making after Ultrasound Diagnosis of Fetal Abnormality," *Reprod Health Matters*, vol. 16, no. sup31, pp. 82–89, Jan. 2008, doi: 10.1016/S0968-8080(08)31372-X.

[7] L. Meng, D. Zhao, Z. Yang, and B. Wang, "Automatic display of fetal brain planes and automatic measurements of fetal brain parameters by transabdominal three-dimensional ultrasound," *Journal of Clinical Ultrasound*, vol. 48, no. 2, pp. 82–88, Feb. 2020, doi: 10.1002/jcu.22762.

[8] A. Ishaq, F. U. M. Ullah, P. Hamandawana, D. J. Cho, and T. S. Chung, "Improved EfficientNet Architecture for Multi-Grade Brain Tumor Detection," *Electronics (Switzerland)*, vol. 14, no. 4, Feb. 2025, doi: 10.3390/electronics14040710.

[9] H. Zhao *et al.*, "Memory-based unsupervised video clinical quality assessment with multi-modality data in fetal ultrasound," *Med Image Anal*, vol. 90, Dec. 2023, doi: 10.1016/j.media.2023.102977.

[10] T. Ciceri, L. Squarcina, A. Giubergia, A. Bertoldo, P. Brambilla, and D. Peruzzo, "Review on deep learning fetal brain segmentation from Magnetic Resonance images," Sep. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.artmed.2023.102608.

[11] M. Alzubaidi, M. Agus, M. Makhlouf, F. Anver, K. Alyafei, and M. Househ, "Large-scale annotation dataset for fetal head biometry in ultrasound images," *Data Brief*, vol. 51, Dec. 2023, doi: 10.1016/j.dib.2023.109708.

[12] S. S. Płotka *et al.*, "Deep learning for estimation of fetal weight throughout the pregnancy from fetal abdominal ultrasound," *Am J Obstet Gynecol MFM*, vol. 5, no. 12, Dec. 2023, doi: 10.1016/j.ajogmf.2023.101182.

[13] J. Lo, A. Lim, M. W. Wagner, B. Ertl-Wagner, and D. Sussman, "Fetal Organ Anomaly Classification Network for Identifying Organ Anomalies in Fetal MRI," *Front Artif Intell*, vol. 5, Mar. 2022, doi: 10.3389/frai.2022.832485.

[14] D. Mavaluru *et al.*, "Advancing fetal ultrasound diagnostics: Innovative methodologies for improved accuracy in detecting down syndrome," *Med Eng Phys*, vol. 126, Apr. 2024, doi: 10.1016/j.medengphy.2024.104132.

[15] N. Hernandez-Cruz *et al.*, "Detection of fetal congenital heart defects on three-vessel view ultrasound videos," *WFUMB Ultrasound Open*, vol. 2, no. 2, Dec. 2024, doi: 10.1016/j.wfumbo.2024.100075.

[16] J. He, L. Yang, B. Liang, S. Li, and C. Xu, "Fetal cardiac ultrasound standard section detection model based on multitask learning and mixed attention mechanism," *Neurocomputing*, vol. 579, Apr. 2024, doi: 10.1016/j.neucom.2024.127443.

[17] R. Ramirez Zegarra *et al.*, "A deep learning approach to identify the fetal head position using transperineal ultrasound during labor," *European Journal of Obstetrics and Gynecology and Reproductive Biology*, vol. 301, pp. 147–153, Oct. 2024, doi: 10.1016/j.ejogrb.2024.08.012.

[18] W. Zhang *et al.*, "A joint brain extraction and image quality assessment framework for fetal brain MRI slices," *Neuroimage*, vol. 290, Apr. 2024, doi: 10.1016/j.neuroimage.2024.120560.

[19] L. Wang, M. Fatemi, and A. Alizad, "Artificial intelligence in fetal brain imaging: Advancements, challenges, and multimodal approaches for biometric and structural analysis," Jun. 01, 2025, *Elsevier Ltd*. doi: 10.1016/j.compbiomed.2025.110312.

[20] P. Ahmed, M. S. U. Yusuf, S. Sarker, A. Chowdhury, A. Bhowmik, and F. Rahman, "A Comparative Study of Different Segmentation and Regression Models for Fetal Head Circumference Measurement," in *2nd International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ic-ETITE58242.2024.10493596.

[21] M. Gong and Q. Fei, "An Improved Light-weight Deep Transfer Learning for Fetal Lung Ultrasound Image Segmentation," in *2024 3rd International Conference on Image Processing and Media Computing, ICIPMC 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 151–156. doi: 10.1109/ICIPMC62364.2024.10586709.

[22] S. Abishek, A. Mathialagan, S. A. Hari Om Swarup, M. Goyal, and J. Mannar Mannan, "An Intelligent Fetal Prognostic System using Machine Learning Techniques," in *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation, IATMSI 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/IATMSI60426.2024.10503455.

[23] M.-L. Zhang and Z.-H. Zhou, "A Review on Multi-Label Learning Algorithms," *IEEE Trans Knowl Data Eng*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014, doi: 10.1109/TKDE.2013.39.

[24] M. A. Talukder, M. A. Layek, M. Kazi, M. A. Uddin, and S. Aryal, "Empowering COVID-19 detection: Optimizing performance through fine-tuned EfficientNet deep learning architecture," *Comput Biol Med*, vol. 168, Jan. 2024, doi: 10.1016/j.compbiomed.2023.107789.

[25] R. Raza *et al.*, "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images," *Eng Appl Artif Intell*, vol. 126, Nov. 2023, doi: 10.1016/j.engappai.2023.106902.

[26] V. Venugopal, N. I. Raj, M. K. Nath, and N. Stephen, "A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images," *Decision Analytics Journal*, vol. 8, Sep. 2023, doi: 10.1016/j.dajour.2023.100278.