

Automated Classification of Fetal and Maternal Structures in Ultrasound Using a Deep Learning Approach

Haifa Ghabri^{1,*}, Mohamed Hamroun², Hedia Bellali³ and Hedi Sakli^{1,4}

¹Communications Systems Lab, École Nationale d'Ingénieurs de Tunis, Tunisia

²XLIM-Lab, UMR CNRS 7252, University of Limoges, Avenue Albert Thomas, Limoges, 87060, France.

³Department of Epidemiology and Statistics, Abderrahmen Mami Hospital, Ariana, Tunisia.

⁴EITA Consulting, 5 rue du chant des oiseaux, 78360 Montesson, France.

Abstract

This study investigates the application of deep learning techniques for the classification of ultrasound images in prenatal and maternal care. The objective was to develop a robust convolutional neural network (CNN) model capable of accurately distinguishing between various anatomical structures including the fetal abdomen, brain, femur, thorax, and maternal cervix. A dataset comprising a diverse range of ultrasound images was used for training and evaluation purposes. The CNN model demonstrated exceptional performance in classification tasks, achieving average precision, recall, and F1 score metrics exceeding 98% across all classes. This indicates the model's capability to effectively identify and differentiate critical features within ultrasound images relevant to fetal and maternal health. The results highlight the potential of deep learning in enhancing diagnostic accuracy and efficiency in prenatal and maternal health monitoring.

Keywords

Ultrasound image classification, multi-class classification, Deep learning, Clinical applications.

1. Introduction

Ultrasound imaging is a critical component of prenatal care, providing invaluable insights into the health and development of the fetus. It is widely used due to its non-invasive nature, safety, and ability to offer real-time visualization of the fetus [1]. Ultrasound scans enable healthcare professionals to monitor fetal growth, assess anatomy, and detect potential anomalies early in the pregnancy. However, the interpretation of ultrasound images requires a high level of expertise and experience, making it a challenging task even for seasoned practitioners. The difficulty in analyzing ultrasound images arises from several factors, including variability in image quality, complexity of fetal anatomy, subtlety of anatomical features, and inter-observer variability. Ultrasound images can vary significantly in quality due to differences in equipment, operator skill, and the physical condition of the patient, which can obscure important anatomical details and make consistent interpretation challenging. The developing fetus undergoes rapid changes, and distinguishing between different anatomical structures requires detailed knowledge and precise image interpretation skills. Many fetal anomalies present subtle signs that can be easily missed without careful and expert examination. Additionally, different clinicians may interpret the same ultrasound images differently, leading to inconsistencies in diagnosis and treatment planning.

Given these challenges, there is a growing interest in applying artificial intelligence (AI) to the field of medical imaging. AI, particularly deep learning, has shown remarkable success in various

WAISS'2024: 1st Euro-Mediterranean Workshop on Artificial Intelligence and Smart Systems, October 15, 2024, Djerba, Tunisia (Co-located with the 17th International Conference on Verification and Evaluation of Computer and Communication Systems (VECoS'2024), October 15-18, 2024, Djerba, Tunisia)

*Corresponding author.

[†]These authors contributed equally.

✉ ghabrihaifa@gmail.com (H. Ghabri); hamrounmohammed@gmail.com (M. Hamroun); fathallahwyssem@gmail.com (H. Bellali)

ORCID 0000-0002-6792-1782 (H. Ghabri); 0000-0002-4618-8673 (M. Hamroun); 0000-0003-0375-3844 (H. Bellali); 0000-0002-7702-6781 (H. Sakli)



© 2025 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

image analysis tasks, from object detection to classification [2], [3], [4], [5]. In the context of fetal ultrasound imaging, AI offers several significant advantages, including consistency and objectivity, efficiency, enhanced accuracy, and scalability. AI systems can provide consistent and objective analysis of ultrasound images, reducing the variability associated with human interpretation. Automated image analysis can significantly speed up the diagnostic process, allowing for quicker decision-making and reducing the workload on healthcare professionals. Advanced AI models can learn to identify subtle patterns and features in ultrasound images, potentially improving the accuracy of fetal anomaly detection. Furthermore, AI systems can be deployed across various healthcare settings, including those with limited access to experienced radiologists, thereby democratizing access to high quality prenatal care. The application of deep learning techniques to ultrasound image analysis has seen significant advancements, particularly in the domain of prenatal and maternal care [6], [7]. This section reviews relevant literature on the use of deep learning for classifying ultrasound images, focusing on key anatomical structures. Several studies have leveraged deep learning techniques to advance the classification of fetal ultrasound images, each contributing unique methodologies and performance metrics.

Zhang et al. [8] introduced a multitask learning-based system for automatic quality assessment of fetal ultrasound images, achieving notable metrics such as Accuracy (0.9431) and AUC (0.9826). Their approach focused on enhancing image quality through precise anatomical identification. Qu et al. [9] proposed a differential-CNN to identify standard fetal brain planes, achieving high accuracy (0.9311) and AUC (0.937). This method effectively discriminated between standard and non-standard views, improving classification accuracy significantly. Montero et al. [10] utilized a GAN-enhanced ResNet model to classify fetal brain images, demonstrating promising results with an accuracy of 0.815 and AUC of 0.867. Their study highlighted the efficacy of GANs in enhancing classification tasks through synthetic image generation. Prieto et al. [11] applied a CNN to classify a diverse dataset of fetal ultrasound images, achieving an accuracy of 0.91.

In this research, we harness the capabilities of the InceptionResNetV2 model, a sophisticated convolutional neural network, to tackle the challenges of fetal ultrasound image classification. This study is distinguished by the integration of a meticulously curated dataset and advanced preprocessing techniques to maximize the performance of the AI model proposed. The dataset, collected in 2020, includes 12,400 2D ultrasound images from 896 pregnant women. These images are categorized into six anatomical regions: maternal cervix, thorax, femur, abdomen, brain, and other.

Our innovative approach incorporates several key steps. We rigorously cleaned the dataset to remove poor quality images and irrelevant information, ensuring a high standard of input data. We precisely cropped the images to focus on the relevant anatomical regions, enhancing the model's ability to learn important features. We also applied the SMOTE techniques to balance the dataset, to ensure an even representation of all categories during training. The core of our methodology is the InceptionResNetV2 model, known for its hybrid architecture combining Inception modules and Residual connections. This design allows the model to efficiently handle high-dimensional data and capture intricate patterns in ultrasound images.

The remainder of this paper is structured as follows. Section II details the proposed model and technique, including model training and parameters. Section III presents the findings of the proposed model. Finally, Section IV concludes with a discussion of our results and potential future work.

2. Methodology

The methodology proposed for fetal ultrasound image classification involved several key steps to ensure robust model training and accurate results. We began by compiling a dataset consisting of 12,400 2D ultrasound images from 896 pregnant women, sourced from hospitals in Barcelona, Spain, in 2020. Each image was meticulously annotated by specialist fetal doctors to label anatomical regions such as the maternal cervix, thorax, femur, abdomen, brain, and other structures. Data preprocessing was crucial to enhance the dataset quality. We conducted rigorous cleaning to remove images with

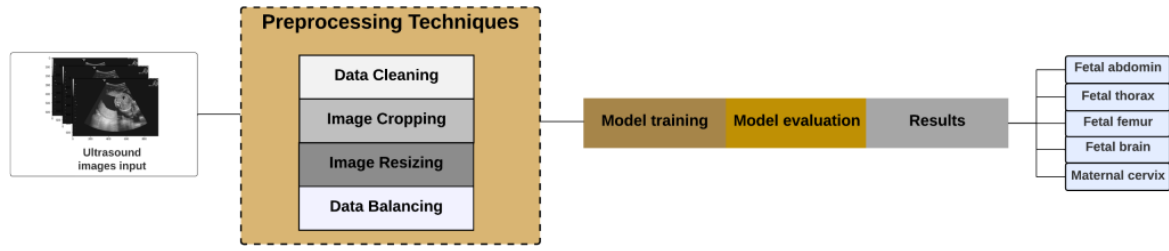


Figure 1: Overview of the proposed methodology.

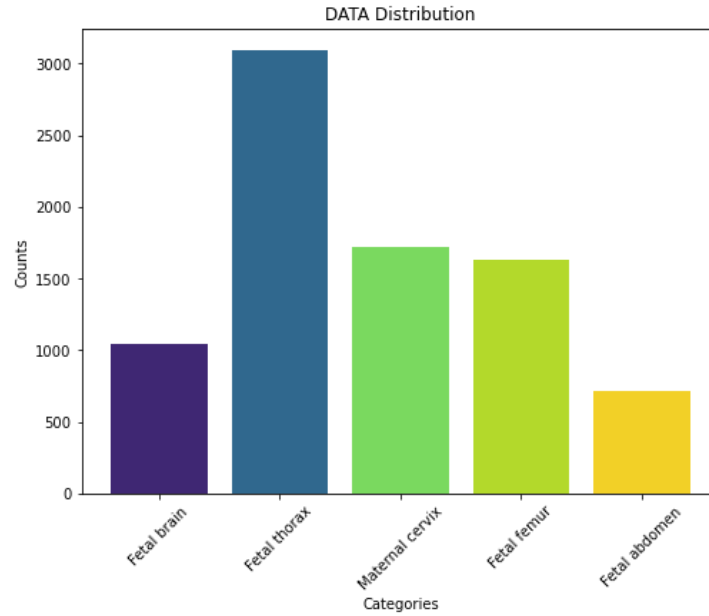


Figure 2: The distribution of the classes.

artifacts or poor quality that could interfere with model training. Additionally, we employed cropping techniques to focus the model's attention on relevant anatomical features while removing extraneous elements from the images. We adopted the InceptionResNetV2 architecture for model development, leveraging its pretrained weights from the ImageNet dataset. This allowed the model to extract complex features from ultrasound images efficiently. Addressing class imbalance was essential, achieved through SMOTE to ensure each anatomical category had sufficient representation during training. Evaluation of model performance included standard metrics such as accuracy and F1-score, providing comprehensive insights into its effectiveness. By integrating these methodologies, our study aims to contribute to advancements in prenatal care by developing a reliable automated system for fetal health assessment through ultrasound imaging. This approach enhances diagnostic capabilities and supports more effective patient care in clinical settings. the proposed methodology is presented in 1.

2.1. Dataset

The dataset used, collected in 2020 from various hospitals in Barcelona, Spain, is comprehensive and diverse, encompassing 12,400 2D ultrasound images from 896 pregnant women[12]. The images are categorized into six distinct anatomical regions: maternal cervix, thorax, femur, abdomen, brain, and other. Each image was meticulously annotated by a specialist fetal doctor, ensuring high quality, reliable labels for training our model. The distribution of the classes is presented in 2.

2.2. Data Preparation

To prepare the dataset for effective model training, we implemented several key data preprocessing steps.

2.2.1. Data Cleaning

In this study on fetal ultrasound image classification, data cleaning played a pivotal role in ensuring the quality and reliability of our dataset. The primary objectives of our data cleaning process were twofold, first, to enhance the clarity and relevance of the ultrasound images. Second, to standardize the dataset for consistent model training. Here's how we approached data cleaning:

- **Artifact Removal:** Ultrasound images often contain artifacts such as shadows, speckles, and machine specific annotations. We applied image processing techniques, such as noise reduction filters and artifact removal algorithms, to eliminate these distractions. By doing so, we ensured that the model focused only on the anatomical structures relevant to fetal health assessment.
- **Normalization and Standardization:** To facilitate effective model training, we normalized the intensity values of the images and standardized their dimensions. This preprocessing step ensured uniformity across the dataset, enabling the model to learn consistent features irrespective of variations in image acquisition parameters.

Preprocessing steps ensure that the model focuses on the most pertinent anatomical regions, thereby improving the accuracy and efficiency of the classification process. The advantages of using preprocessing in our research are significant. It enhances the model's ability to concentrate on the essential features of each anatomical region, such as the maternal cervix, thorax, femur, abdomen, and brain parts. This focused learning improves the model's capability to differentiate between similar anatomical structures and identify subtle anomalies. This consistency is crucial for achieving reliable performance across different ultrasound machines and varying image acquisition conditions.

2.2.2. Data Balancing

In this research, we employed Synthetic Minority Over-sampling Technique (SMOTE) [13] to address the challenge of class imbalance in the fetal ultrasound image dataset. SMOTE is a powerful data augmentation technique specifically designed to balance class distribution by generating synthetic examples for minority classes. Unlike traditional oversampling methods that simply duplicate existing minority class samples, SMOTE creates new synthetic instances by interpolating between existing samples. This approach enhances the diversity of the training set and helps the model to learn more generalized features.

The use of SMOTE in our dataset, which consists of 12,400 2D ultrasound images categorized into five anatomical regions, is particularly advantageous. Fetal ultrasound images often exhibit significant class imbalances, with some anatomical regions being underrepresented compared to others. This imbalance can lead to biased model training, where the model becomes more accurate in predicting the majority classes while underperforming on the minority ones. By applying SMOTE, we effectively mitigated this issue, ensuring that each class is adequately represented in the training process.

2.3. Model Architecture: InceptionResNetV2

In this research, the choice of model played an essential role in achieving accurate and reliable results. We opted for the InceptionResNetV2 architecture [14], a state of the art convolutional neural network (CNN), renowned for its effectiveness in handling complex image recognition tasks.

In our implementation, we leveraged the pretrained weights of InceptionResNetV2 trained on the ImageNet dataset. Transfer learning from ImageNet provides a significant advantage by initializing the model with weights that have already learned general features from a vast and diverse set of natural

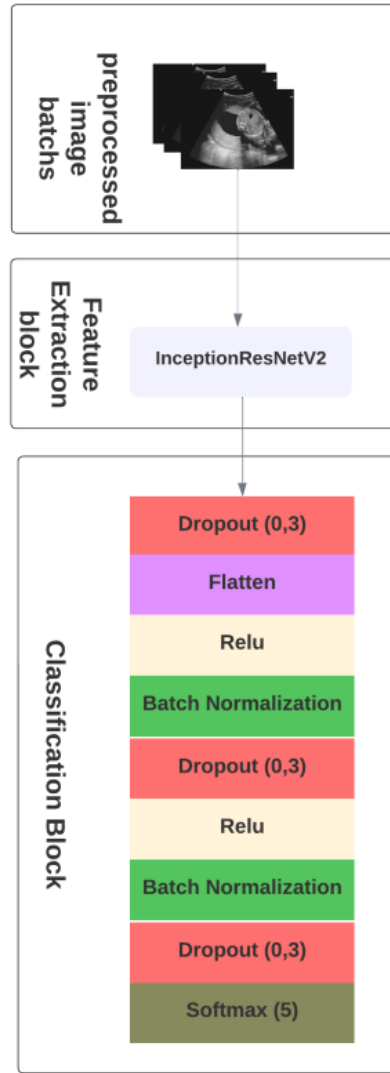


Figure 3: The overall architecture of the model proposed.

images. Fine-tuning the model on our specific fetal ultrasound dataset further tailored its parameters to better recognize anatomical structures such as the maternal cervix, thorax, femur, abdomen, and brainregions to fetal health assessment. After importing the pretrained InceptionResNetV2 without its fully connected layers, we introduced dropout layers strategically placed after the convolutional base and dense layers. These dropout layers mitigate overfitting by randomly deactivating neurons during training, promoting better generalization of the model. Following the dropout layers, we incorporated a flatten layer to reshape the output into a 1D vector, preparing it for input into subsequent dense layers. These dense layers, equipped with ReLU activation functions, facilitate the learning of complex, nonlinear relationships within the data, crucial for distinguishing between different structures in fetal ultrasound images. To stabilize and accelerate training, batch normalization layers were added after each dense layer, normalizing the activations and improving the model's convergence speed and overall performance. This tailored approach not only leverages the powerful feature extraction capabilities of InceptionResNetV2 but also optimizes it for the nuanced requirements of medical image classification, ultimately advancing automated diagnostic tools for prenatal care.

In our study, we optimized various hyperparameters to enhance the performance of our model for fetal ultrasound image classification. The Adam optimizer was selected due to its adaptive learning

Table 1
Hyperparameters used in the proposed approach.

Hyperparameters	Optimized value
Optimizer	Adam
Learning rate	0.001
Loss function	Categorical cross-entropy
Train Validate and Test ratio	64:16:20
Batch size	64
Epochs	50

rate capabilities, which help in efficiently handling the sparse gradients encountered in our dataset. We set the learning rate at 0.001, striking a balance between training speed and convergence stability. The categorical cross entropy loss function was employed to handle the multiclass classification task effectively. Our dataset was divided into training, validation, and test sets in a ratio of 64:16:20, ensuring a robust evaluation of our model’s generalizability. A batch size of 64 was chosen to make the training process manageable while maintaining computational efficiency. Finally, the model was trained for 50 epochs, allowing sufficient iterations to learn complex patterns in the data without overfitting. These carefully selected hyperparameters contributed significantly to the superior performance metrics achieved in our study.

3. Results and discussion

The proposed approach yielded promising results, demonstrating significant improvements in the accuracy and robustness of fetal ultrasound image classification compared to existing methods. The InceptionResNetV2 model achieved high classification accuracy across all five categories, with notable performance in distinguishing between anatomically similar regions such as the thorax and abdomen.

The training process spanned 50 epochs, during which both training and validation datasets were utilized to optimize the model. Initially, the model achieved a validation loss of 0.8232 and an accuracy of 95.23% in the first epoch. This initial validation performance provided a benchmark for subsequent epochs.

As training progressed, there was a consistent improvement in both training and validation metrics, indicating that the model effectively learned to generalize to unseen data. The validation accuracy regularly increased, reaching a peak of 99.60% by the 31st epoch. Notably, the validation loss continued to decrease, suggesting that the model’s predictions became more precise and aligned with ground truth labels. Despite the model’s impressive performance, there were instances where the training accuracy exceeded the validation accuracy slightly, suggesting some degree of over-fitting. However, the consistent decrease in validation loss and increase in validation accuracy throughout most epochs indicate that the model learned to generalize well to new data, mitigating overfitting to a considerable extent.

The model achieved a peak validation accuracy of 99.60%, indicating that it correctly classified the majority of samples in the validation set. The validation loss decreased from 0.8232 in the first epoch to as low as 0.0142 by the last epoch, highlighting the model’s improved predictive accuracy and consistency.

Figure 4 presents the evaluation of performance metrics throughout the training and validation phases, focusing on loss and accuracy.

Throughout the training, both precision and recall metrics consistently improved. Precision measures the model’s ability to correctly identify positive instances among all predicted positive instances, while recall measures the model’s ability to correctly identify positive instances among all actual positive instances. By the end of training, precision and recall values were consistently above 99%, indicating high confidence in the model’s predictions.

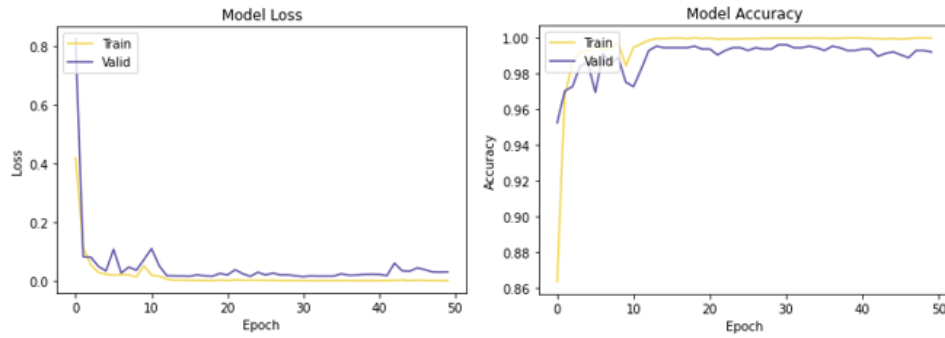


Figure 4: The evaluation of performance metrics throughout the training and validation phases for loss and accuracy metrics.

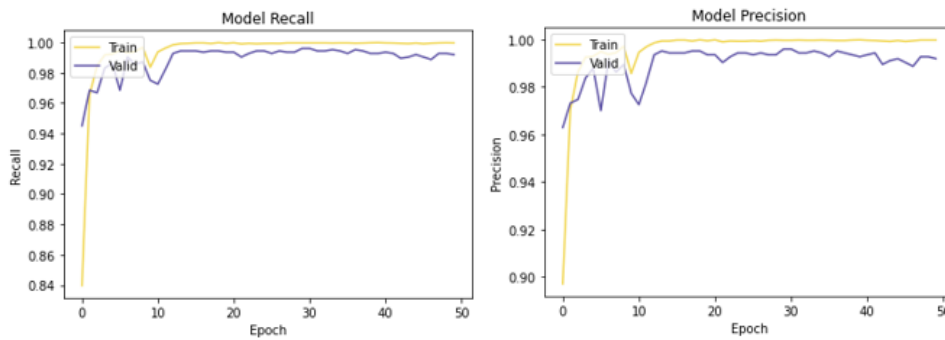


Figure 5: The evaluation of performance metrics throughout the training and validation phases for Recall and Precision metrics.

Figure 5 displays the evaluation of performance metrics throughout the training and validation phases, focusing on Recall and Precision.

To evaluate the performance of the model used on the test dataset, we employed a confusion matrix to analyze the classification accuracy across five distinct classes: fetal abdomen, fetal brain, fetal femur, fetal thorax, and maternal cervix.

The confusion matrix revealed high precision and recall values for each class, indicating the model's robust ability to correctly identify and distinguish between these anatomical structures.

Figure 6 depicts a confusion matrix, a common tool used to visualize the performance of a classification model across multiple classes.

Our approach demonstrates a superior accuracy of 0.9909 compared to previous studies, showcasing a significant advancement in the classification of fetal ultra-sound images. The precision and recall of our model also surpass those of the other models, indicating a more balanced and reliable performance across various metrics. For example, while Zhang et al. achieved a high AUC of 0.9826, our model excels in overall accuracy, suggesting a robust capability in handling diverse and challenging ultrasound images. Qu et al.'s differential-CNN also performed well, but our higher precision and recall values highlight the efficacy of our method in accurately identifying fetal anatomical structures. Montero et al.'s use of GANs, although innovative, resulted in lower accuracy, which our model improves upon significantly.

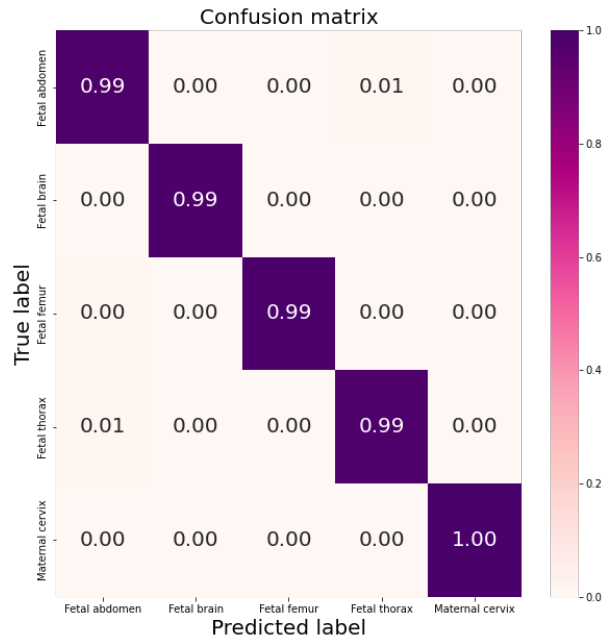


Figure 6: Confusion matrix for classification accuracy.

Table 2

Comparison of models for fetal ultrasound image classification.

Study	Classes	Model	Accruacy	Precision	Recall
Zhang et al.	1	CNN	0.9431	0.9463	0.9241
Qu et al.	6	DCNN	0.9311	0.9262	0.9239
Montero et al.	2	GAN	0.815	-	-
Our work	5	InceptionResNetV2	0.9909	0.9908	0.9918

4. Conclusion

Overall, the model exhibits excellent performance across all evaluated metrics. It shows high precision and recall for each class, typically above 98%, indicating accurate predictions and effective identification of relevant instances. The F1 score also reflect a balanced trade off between precision and recall, consistently above 98%. With an overall accuracy of 99.09%, the model correctly predicts class labels for the majority of instances, demonstrating its effectiveness in distinguishing between different classes. The support (number of instances) for each class is well-distributed, which helps ensure the model learns effectively without bias towards any specific class. In summary, the model's strong performance across accuracy, precision, recall, and F1 scores indicates its robustness and reliability in classifying instances into their respective classes. It is well suited for practical applications where accurate classification is essential. For future work, several avenues could be explored to potentially enhance the model's performance and address broader considerations, experimenting with different neural network architectures or exploring more advanced models (e.g., deeper networks, attention mechanisms) could potentially capture more intricate patterns in the data and further boost performance. Thoroughly validating the model's performance in real-world settings and monitoring its performance post-deployment to ensure consistency and reliability in diverse conditions.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] M. C. Fiorentino, F. P. Villani, M. Di Cosmo, E. Frontoni, and S. Moccia, “A review on deep-learning algorithms for fetal ultrasound image analysis”, *Med. Image Anal.*, vol. 83, p. 102629, Jan. 2023, doi: 10.1016/j.media.2022.102629.
- [2] M. Sakli, C. Essid, B. Ben Salah, and H. Sakli, “Deep Learning-Based Multi-Stage Analysis for Accurate Skin Cancer Diagnosis using a Lightweight CNN Architecture”, in 2023 International Conference on Innovations in Intelligent Systems and Applications (INISTA), Sep. 2023, pp. 1–6. doi: 10.1109/INISTA59065.2023.10310615.
- [3] S. Abdelbaki et al., “Improving diagnosis accuracy with an intelligent image retrieval system for lung pathologies detection: a features extractor approach”, *Sci. Rep.*, vol. 13, Oct. 2023, doi: 10.1038/s41598-023-42366-w.
- [4] M. Sakli, C. Essid, B. B. Salah, and H. Sakli, “Lightweight CNN Towards Skin Lesions Automated Diagnosis In Dermoscopic Images”, in 2023 International Conference on Innovations in Intelligent Systems and Applications (INISTA), Sep. 2023, pp. 1–6. doi: 10.1109/INISTA59065.2023.10310480.
- [5] N. Benameur, R. Mahmoudi, N. Benameur, and R. Mahmoudi, “Deep Learning in Medical Imaging”, IntechOpen, 2023. doi: 10.5772/intechopen.111686.
- [6] H. Ghabri, W. Fathallah, H. Sakli, and M. N. Abdelkarim, “Enhancing Maternofetal Ultrasound Images Toward Boosting Classification Performance on a Diverse and Com-prehensive Data,” in 2023 International Conference on Innovations in Intelligent Sys-tems and Applications (INISTA), Sep. 2023, pp. 1–6. doi: 10.1109/INISTA59065.2023.10310548.
- [7] T. G. Day, B. Kainz, J. Hajnal, R. Razavi, and J. M. Simpson, “Artificial intelligence, fetal echocardiography and congenital heart disease,” *Prenat. Diagn.*, vol. 41, no. 6, pp. 733–742, 2021, doi: 10.1002/pd.5892.
- [8] B. Zhang, H. Liu, H. Luo, and K. Li, “Automatic quality assessment for 2D fetal sonographic standard plane based on multitask learning,” *Medicine (Baltimore)*, vol. 100, no. 4, p. e24427, Jan. 2021, doi: 10.1097/MD.00000000000024427.
- [9] R. Qu, G. Xu, C. Ding, W. Jia, and M. Sun, “Standard Plane Identification in Fetal Brain Ultrasound Scans Using a Differential Convolutional Neural Network,” *IEEE Access*, vol. 8, pp. 83821–83830, 2020, doi: 10.1109/ACCESS.2020.2991845.
- [10] A. Montero, E. Bonet-Carne, and X. P. Burgos-Artizzu, “Generative Adversarial Networks to Improve Fetal Brain Fine-Grained Plane Classification,” *Sensors*, vol. 21, no. 23, Art. no. 23, Jan. 2021, doi: 10.3390/s21237975.
- [11] J. C. Prieto et al., “An automated framework for image classification and segmentation of fetal ultrasound images for gestational age estimation,” in *Medical Imaging 2021: Image Processing*, B. A. Landman and I. Išgum, Eds., Online Only, United States: SPIE, Feb. 2021, p. 55. doi: 10.1117/12.2582243.
- [12] X. P. Burgos-Artizzu et al., “Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes,” *Sci. Rep.*, vol. 10, no. 1, p. 10200, Dec. 2020, doi: 10.1038/s41598-020-67076-5.
- [13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
- [14] F. Baldassarre, D. G. Morín, and L. Rodés-Guirao, “Deep Koalarization: Image Colorization using CNNs and InceptionResNetv2,” *arXiv.org*. Accessed: Mar. 02, 2023. [Online]. Available: <https://arxiv.org/abs/1712.03400v1>