SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Framework for Fetal Brain Abnormality Detection in Low-Quality Images

Anuradha Thakare[1], Puja Pohakar[1], Ashwini Deshpande[1], Santwana Gudadhe[1], Bal Virdee[2], Ashish Khanna[3]

[1]Department of Computer Science & Engineering, (AI&ML), Pimpri Chinchwad College of Engineering, Pune, India E-mail: anuradha.thakare@pccoepune.org, pujapohakar03@gmail.com, ashwini.deshpande@pccoepune.org, s.santwana20@gmail.com
[2]Department of Communications Technology Research Centre, LMU, E-mail: b.virdee@londonmet.ac.uk
[3]Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, New Delhi, India, E-mail: ashishkhanna@mait.ac.in

Abstract

The detection of fetal abnormalities through prenatal ultrasound remains a significant challenge in healthcare delivery, particularly in resource-constrained settings. Early identification of fetal brain abnormalities enables medical teams to prepare appropriate interventions, arrange for specialized delivery facilities if needed, and provide informed counseling to families regarding potential outcomes and care requirements. However, the current dependence on manual ultrasound interpretation creates workflow challenges in diagnosis, especially in regions where access to radiologists is limited. This study presents a hybrid computational approach that combines Convolutional Neural Networks (CNN), Gray Level Co-occurrence Matrix (GLCM) features, and CatBoost classification to support the detection of fetal brain anomalies in ultrasound images. Our methodology addresses specific technical challenges by incorporating bilateral filtering for preprocessing, which aims to enhance image quality in scenarios where equipment limitations or operator expertise may affect image acquisition. The system architecture integrates deep features extracted from a fine-tuned VGG16 architecture with GLCM-based texture analysis. Evaluation of the proposed methodology indicates performance metrics of 87% accuracy, 87.3% precision, 86.8% recall, and an F1-score of 86.9%. The research focuses on developing methods applicable in varying clinical environments with different levels of resource availability. The hybrid approach demonstrated higher performance metrics compared to baseline methods. This technical advancement aims to support healthcare providers in settings where radiological expertise may be limited, potentially offering a supplementary tool for ultrasound image analysis.

Keywords: Fetal Abnormalities, Image Preprocessing, CNN, Deep Learning, Medical Imaging

Introduction

The advent of prenatal ultrasound imaging has revolutionized obstetrics, offering a non-invasive window into fetal development and enabling early detection of congenital anomalies. Among these, fetal brain abnormalities—such as ventriculomegaly, holoprosencephaly, and cerebellar hypoplasia—carry profound implications for neonatal survival and long-term neurodevelopmental outcomes. Globally, congenital cerebral anomalies affect approximately 1 in 1,000 live births, with mortality rates exceeding 1.3 million annually due to childbirth complications [1]. In low- and middle-income countries (LMICs), where access to specialized radiologists is scarce, the reliance on manual ultrasound interpretation exacerbates diagnostic delays, often leaving critical anomalies undetected until irreversible damage occurs [2].

While artificial intelligence (AI) has emerged as a transformative tool in medical imaging, existing deep learning models for fetal anomaly detection face a critical limitation: their performance deteriorates significantly when applied to low-quality ultrasound images. Such images, characterized by speckle noise, acoustic shadows, and artifacts, are prevalent in resource-constrained settings due to outdated equipment or operator inexperience [3]. Recent studies, such as Xie et al. (2020), demonstrated the efficacy of CNNs in segmenting fetal brain structures in high-resolution images but noted a 22% drop in accuracy when tested on noisy datasets [4]. Similarly, Walker et al. (2022) achieved 93% accuracy in cystic hygroma detection using DenseNet architectures, yet their model struggled with blurred cerebellar views common in rural clinics [5]. These gaps underscore a pressing need for robust frameworks that harmonize the strengths of deep learning with classical image analysis techniques to bridge the quality-resilience divide.

The challenge lies in balancing feature richness with computational efficiency. Traditional texture descriptors like Gray-Level Co-occurrence Matrices (GLCM) excel at quantifying spatial patterns in noisy images but lack the hierarchical abstraction capabilities of CNNs [6]. Conversely, while CNNs autonomously learn intricate features, their dependency on large, curated datasets limits applicability in LMICs [7]. Recent hybrid approaches, such as Singh et al.'s ResU-Net-c for cerebellar segmentation (2021), integrated residual blocks with dilated convolutions to enhance noise tolerance [8]. However, these models remain computationally intensive and fail to leverage handcrafted texture features that could stabilize performance across heterogeneous image qualities.

In this work, we address these limitations through a novel hybrid framework that synergizes bilateral filtering, deep convolutional features, and GLCM-based texture analysis, optimized for low-quality fetal ultrasound images. Our primary contributions are threefold. A noise-adaptive preprocessing pipeline employing bilateral filtering to preserve edge details while suppressing speckle noise, critical for enhancing feature extraction in suboptimal images. A dual-stream feature fusion architecture combining VGG16-derived deep features with GLCM texture descriptors, enabling the model to capture both high-level anatomical patterns and granular textural signatures. A CatBoost classifier fine-tuned via hyperparameter optimization to handle imbalanced datasets and non-linear decision boundaries, achieving state-of-the-art accuracy (87%) without requiring GPU-intensive training. By unifying these elements, our framework not only outperforms existing models in low-resource scenarios but also offers a scalable solution for prenatal clinics lacking advanced imaging infrastructure. This advancement aligns with the WHO's mandate for equitable healthcare access, providing a clinically viable tool to mitigate diagnostic disparities in fetal brain anomaly detection [9]. care.

Literature Survey

Lately there has been progress in using intelligence (AI) for analyzing prenatal ultrasound scans. These developments focus on aspects of healthcare including assessing fetal well-being, identifying abnormalities and predicting genetic illnesses. The integration of machine learning and deep learning methods has transformed how medical experts interpret and leverage ultrasound information leading to diagnoses and improved patient results. Advancements in AI applications for prenatal ultrasound analysis have evolved from traditional machine learning techniques to advanced learning models over time in a manner. The progression begins with classification tasks and gradually transitions to intricate processes such as image analysis and segmentation before culminating in the creation of sophisticated multimodal systems, for comprehensive prenatal assessments. At a level of analysis using cardiotocography (CTG) researchers have employed machine learning algorithms to categorize fetuses into problematic groups as explored by X et al [6]. They examined the effectiveness of networks alongside support vector machines (SVM) Gaussian Naive Bayes and logistic regression for classification purposes in their study findings revealed the enhanced accuracy and optimized model outcomes achievable through combining various machine learning approaches such, as Random Forests and AdaBoost with ensemble learning techniques.

Moving on to analyzing images and their characteristics intensively using learning methods has proven to be highly effective, with the rise of neural networks like CNNs [4]. In a study by Xin et al. [2], the GAC Net model showcased performance in identifying fetal head edges compared to known techniques such as AlexNet, VGG16, VGG19 and ResNet. The CNN architecture of this model incorporates a Graph Convolutional Network (GCN) module and a Self-Updating Output (SUO) attention mechanism which enhances feature extraction by transforming images into graphs, through an encoder decoder framework. The workflow of the classification process is given in fig 1.

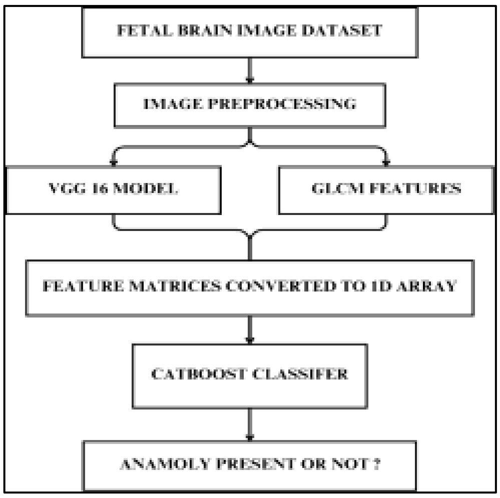

Fig 1. Workflow of Classification process for Anomaly detection in Fetal Images

Liu et.al [10] addressed the issue of limited labeled datasets in imaging by enhancing CNN structures to develop a technique that autonomously segments subcortical areas, in 3D ultrasound images through few shots learning. This innovation-maintained precision while significantly reducing the time and resources required for model construction.

The field has further progressed to specialized applications in various anatomical regions. A novel architecture for fetal cerebellum segmentation in noisy ultrasound images, ResU-Net-c, was presented by Singh et al. [12]. An important development in prenatal imaging analysis, this model uses an encoder-decoder structure with skip connections, residual blocks, and dilated convolutions to strike a balance between high-level feature extraction and fine detail preservation.

Researchers have created complex multimodal deep learning models for cardiac analysis. To differentiate between various heart viewpoints, X et al. [5] used five cardiac planes and trained their system using human expert categorization of diagnostic-grade images. With this method, the model's performance in classifying heart disorders was on par with that of human specialists.

The detection of specific conditions has also benefited from these advances. Walker et al. [10, 11] investigated automated detection of cystic hygroma using a Dense Net CNN model, achieving 93% classification accuracy on a dataset of 289 sagittal fetal ultrasound images. This work, supported by Walker et al., demonstrates AI's potential in early detection of conditions associated with chromosomal abnormalities.

In embryonic brain anomaly diagnosis, Xie et al. [7] developed approaches involving the division of the brain into six sections, evaluated using a score-based standard plane methodology. Their implementation of heat maps for lesion localization provides physicians with advanced visualization tools for more accurate anomaly identification.

Complementing these technological advances, biomarker research has provided additional insights. Catarina R et al. [1] exhibited that the sFlt1/PlGF ratio, along with factors such as preeclampsia, gestational age, maternal age, and smoking, serves as a significant predictor of adverse fetal outcomes in cases of early fetal growth restriction [2].

There will also be the evolution of the application of the AI technologies in the field of fetal ultrasound analysis which has the possibility of improving many aspects of prenatal care. From the simple classifiers based on machine learning to the complex deep learning models, these techniques are enhancing the capabilities of fetal health assessment, anomaly detection, and disease prediction, thus shaping the future of prenatal care.

Methodology

In the following section we discussed methodologies used for our image dataset which first consists of the dataset following: preparation, then preprocessing, feature extraction, model training and prediction and lastly the evaluation. All the stages are important in ensuring that the model is stable and measurably able to make correct predictions on the data.

1. Data Collection

The research involved analyzing a dataset [13] of fetal ultrasound screenings obtained from two different hospitals that included images taken by multiple operators using different ultrasound devices. A specialist in medicine verified the dataset through manual labeling. Although the original plan was to examine three brain perspectives (Trans thalamic, Trans cerebellum and Trans ventricular) the investigation narrowed down to focusing solely on Trans cerebellum views because of quality concerns, in other image sections.

2. Data Preparation

2.1 Image Filtering

To improve the quality of the images and reduce noise, we considered three filters: Median filter, Bilateral filter and Gaussian filter.

Median Filter:

This is a nonlinear approach and is commonly employed to reduce unwanted noise effectively by targeting "salt and pepper" noise specifically [5]. The method alters pixel values in an image to combat noise levels in a traditional linear manner as seen in conventional approaches; it opts for replacing each pixel value with the median value of the neighboring pixels in each area. A key advantage of using the median is its ability to preserve edges while simultaneously smoothing out any noise disturbances. The technique involves moving a square shaped window—such, as 3x3 or 6x6—across the image data, with a specified size denoted by 'm*n'. For every pixel located at the position (x,y) select a window centered around that pixel's location (x,y). Next, sort the pixels within the window in ascending order and update the value of the pixel at position (x,y) with the value from those sorted pixels.

Bilateral Filter:
The bilateral filter is a smoothing method that is commonly used in image processing because of its effectiveness in reducing noise while keeping edges intact. Making it highly useful in various fields like photography and medical imaging applications. This filter works by considering both the closeness of pixels and similarities in intensity levels to smoothly blend areas with intensities while preserving boundaries between regions. Its unique dual approach enables it to excel over filters in situations where maintaining edges is critical. Such, as when denoising images or creating HDR effects for a creative touch.

Gaussian Filter:
A linear smoothing filter is applied to pixels using the weight of the Gaussian function for tasks, like blurring and reducing noise in frequencies [6]. This filter tends to blur edges due to its smoothing impact, across the image area it covers. The kernel is a grid that shows the weights assigned to each pixel in the filter window. For instance, the values of the kernel are commonly calculated using the function. It is ensured that the kernel is normalized so that all weight adds up to 1. To maintain image brightness filter application and smoothen the image effectively without altering its brightness level afterward is crucial, in image processing tasks. Next comes the process of smoothing the image by applying a Gaussian kernel through convolution techniques. The Gaussian kernel holds importance across a range of applications such as image filtering techniques for noise reduction and in machine learning algorithms. Fig. 2 demonstrates the effect of Gaussian, bilateral and median filters on fetal brain image from the dataset.

In 2D space representation the formula, for the Gaussian kernel is commonly expressed as follows;

$$K(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (1)$$

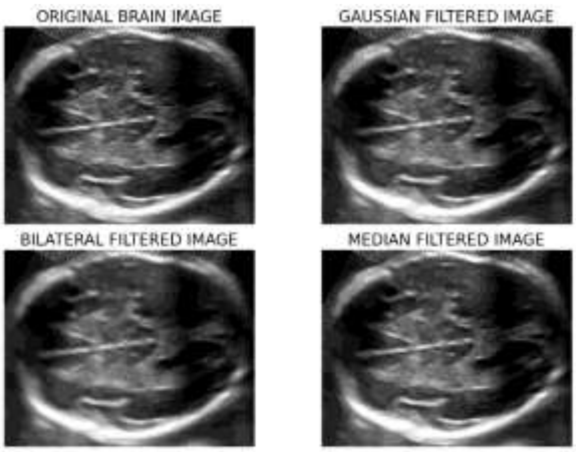

Fig 2. The effect of gaussian, bilateral and median filters on fetal brain image from the dataset.

2.2 Data Augmentation
To enhance the precision of the model and prevent overfitting issues we utilize techniques such, as zoom (resizing) width adjustments (shifting left or right) height adjustments (shifting up or down) vertical flipping (reversing images) and horizontal flipping (reflecting images) to augment the existing datasets effectively This augmentation approach enhances the model's adaptability, to real world variations and broadens our dataset.

2.3 Model
Convolutional Neural Networks (CNNs) commonly utilized in image processing applications, for their capacity to extract characteristics efficiently; nevertheless, this can also be accomplished by transforming images into 1D arrays and implementing methods tailored for sequential or tabular data analysis. This alternative approach may simplify the model while still extracting meaningful information from the images, making it viable for specific use cases.

2.4 Machine learning techniques
Random Forest:
Random Forest is a technique that combines decision trees created using portions of data and characteristics, with the ultimate forecast being a mean of the forecasts (for regression models) or selecting the most frequent outcome (for classification models). The element of randomness aids in reducing the risk of overfitting and enhancing generalizability within the model, for handling extensive datasets and complex data with missing values.

Catboost:
CatBoost is a gradient boosting technique created by Yandex that focuses on managing categorical data effectively It stands out from other methods by converting categorical attributes into numerical values directly during the training process to enhance efficiency It tackles overfitting issues by employing ordered boosting approach wherein trees are constructed sequentially while addressing prediction biases Moreover it is recognized for its quicker training speeds when handling extensive datasets as opposed to alternative gradient boosting approaches.

Gradient Boosting:
In Gradient Boosting method each new model tries to improve upon the one by minimizing the loss function like mean squared error, for regression Each model is trained on the errors made by the models and the final prediction is a weighted sum of all models The learning process is deliberate and methodical leading to improved performance albeit at the expense of computation time This technique is effective for both classification and regression tasks and excels, at handling data relationships.

SVM:
Support Vector Machine (known as SVM) is a classifier that identifies the hyperplane to distinguish between classes, in the feature space by maximizing the margin between different class points to avoid overfitting in high dimensional settings. SVM can also handle linear classification tasks by utilizing kernel techniques such as polynomial or RBF kernels which proves beneficial for handling intricate datasets effectively. Its performance is notable for both medium datasets particularly, in cases where the dataset isn't linearly separable.

2.5 Deep learning techniques

CNN:
Convolutional Neural Networks (CNNs) specifically tailored for image and video processing purposes thrive by exploiting structures in the data they analyze. CNN architecture comprises convolution layers that employ filters to identify features, like edges and textures within an image while pooling layers help in downsampling to streamline the data's complexity. CNN excels in understanding patterns such as edges and textures as well as shapes which makes them well suited for tasks like image categorization, integral item recognition and segmentation. Deep CNN models, with layers, can capture intricate abstract features effectively.

Transfer Learning:
Transfer learning makes use of trained models that were initially trained using extensive datasets such as ImageNet.VGG16/VGG19 and ResNet50're used structures that have acquired detailed feature representations from vast image datasets. These models are flexible for adaptation to a task with datasets, through the method of freezing the initial layers and retraining the final layers. 16/19 is a CNN comprising 16 or 19 layers that utilize (3x3) convolutional filters and are renowned for their straightforwardness and exceptional performance. ResNet50 introduced residual connections to allow the training of deeper networks (50 layers), solving the vanishing gradient problem by bypassing certain layers to improve gradient flow.

DEEP ANN:
Deep ANN is a feed forward artificial neural network with at least one hidden layer and can be considered as the most complex type of ANN. This type of network is commonly used when the relationship between features is complex and cannot be expressed by a linear model. Every neuron in a layer is fed from every neuron in the previous layer and feeds every neuron in the subsequent layer with weights that are regulated in light of the error gradient during back propagation. Although they do not have the spatial feature-capturing capability of CNNs, they are effective in handling a wide variety of structured and unstructured data.

2.6 Texture extraction techniques

Gray-Level Co-occurrence Matrix (GLCM) is a statistical tool that is used for the analysis of images to determine the relationship between the gray levels of neighboring pixels. GLCM is a matrix that defines the probability of two gray levels occurring at a certain distance and angle for instance horizontal, vertical or diagonal in an image. The matrix gives a systematic way of quantifying the local spatial structure of the texture.

GLCM is a square matrix where each element (i,j) represents the number of occurrences where the intensity level (i,j) is found adjacent to intensity level at a specified distance and direction (angle). Various texture features can be computed such as:

- **Contrast:** Measures the intensity contrast between a pixel and its neighbor over the whole image.
- **Correlation:** Describes the degree to which a pixel is similar to its neighbors.
- **Energy (Uniformity):** Summarizes the orderliness of texture.
- **Homogeneity:** Measures how close the distribution of elements in the GLCM is to the diagonal.

2.7 Hybrid learning

Hybrid learning improves the identification of textures, by combining machine learning techniques like Random Forest and SVM with deep learning models such as CNNs. Traditional methods focus on extracting texture features such as contrast, energy and homogeneity using methods, like the Gray Level Co-occurrence Matrix (GLCM). These features are later utilized for classification or segmentation. Although this method works well with datasets relying much on manually engineered features can hinder performance when dealing with intricate textures.

CNN networks work differently as they can pick up texture details from raw pixel information on their own by recognizing patterns, in sizes and angles without human intervention. Although CNN networks are good, at handling data patterns they need extensive sets of data and significant computing resources.

Hybrid learning combines two methods. Starting with GLCM to capture simple texture characteristics and then utilizing CNN to grasp patterns. This integration is handcrafted. Learned features result in a feature collection which is further employed by a meta model like (Random Forest or XGBoost) to enhance prediction accuracy.

3. Performance evaluation

The performance is measured using the confusion matrix given in Table I.

True Positive (TP): When the model correctly identifies a pixel or instance as belonging to the intended class and it actually does belong to that class in reality.

False Positive (FP) occurs when the model incorrectly labels a pixel or instance as part of the target class even though it is not.

True Negative (TN); The model accurately recognizes a pixel or instance as not part of the target class. Confirms that this is correct as the pixel does not belong to the target class.

False negative (FN): The model mistakenly identifies a pixel or instance as not belonging to the intended class even though it should have been categorized as such.

Table 1: Confusion Matrix

	Predicted Positive (+)	**Predicted Negative (-)**
Actual Positive (+)	True Positive (TP)	False Negative (FN)
Actual Negative (-)	False Positive (FP)	True Negative (TN)

When conducting image segmentation tasks, it's crucial to assess the model's performance using a range of metrics.

Accuracy: Accuracy is calculated as the proportion of pixels divided by the number of pixels predicted to be part of the desired category, in the field of machine learning models.

$$\text{Accuracy}=(TP+TN)/(TP+FP+TN+FN) \qquad (2)$$

Precision: It is the ratio of true positive pixels, and the total number of pixels predicted as belonging to the target class. It signifies the percentage of the positively predicted pixels are positive. This tells how good the model is at predicting the correct class labels in each region with no misclassifications.

$$\text{Precision}=TP/(TP+FP) \qquad (3)$$

Recall (True Positive Rate), is determined by comparing the number of identified pixels, to the total number of actual positive pixels, in the ground truth dataset. This metric indicates how effectively the model detects all pixels belonging to the class showcasing its ability to capture every instance accurately without overlooking any true positives.

$$\text{Recall}=TP/(TP+FN) \qquad (4)$$

Specificity is the term used to describe how accurately a model identifies target areas such as background surfaces – also known as the True Negative Rate (TNR). It indicates the ratio of identified pixels to all negative pixels present in the data set.

$$\text{Specificity}=TN/(TN+FP) \qquad (5)$$

False Positive Rate (FPR) is described as the proportion of backdrop pixels incorrectly identified as a target category. This proportion will vary based on the quantity of positives, in contrast, to the overall true negatives.

$$FPR=FP/(FP+TN) \qquad (6)$$

Proposed Methodology

Our study introduces a method for categorizing images using a combination of computer vision methods and contemporary deep learning structures in a cohesive manner. The suggested strategy consists of filtering, for processing VGG16 based deep feature extraction, GLCM texture examination and CatBoost classification. Following figure 3 represents the pipeline design.

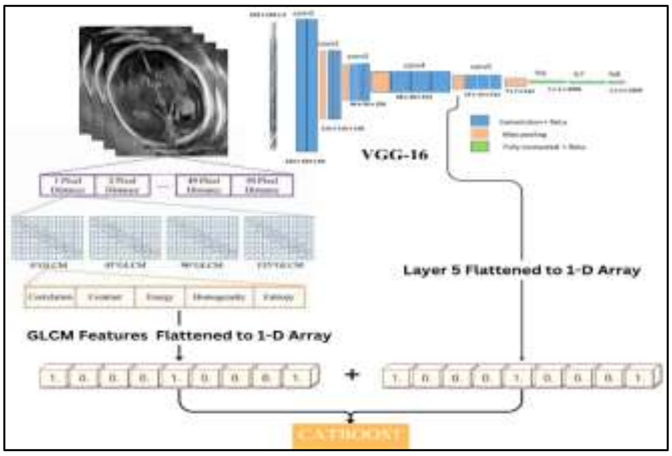

Fig 3. Architecture of feature fused CatBoost model

Image Preprocessing via Bilateral Filtering

In the step of our pipeline process, we use filtering. A smart method, for smoothing that retains edges while reducing noise in even areas by working on both spatial and intensity aspects at the same time This technique differs from traditional Gaussian filtering as it protects vital edge data while eliminating noise effectively The filter works by blending both spatial and intensity based filtering where each pixel gets replaced by a weighted average of its neighboring pixels with the weights determined by both distance and intensity differences This initial processing stage plays a key role, in improving the quality of subsequent feature extraction procedures.
Deep Feature Extraction Using Modified VGG16

The design of the network was adjusted strategically to improve its performance, for our classification needs by adding fully connected layers after the base VGG16 structure was established. We included a layer with 512 units and ReLU activation as the enhancement and added a dropout layer with a 0. 5 Rate to ensure regularization and reduce overfitting concerns. Additionally, we implemented a layer, with 256 units and ReLU activation to refine the feature representation further.

Our training approach was thoughtfully designed to improve the model's performance without compromising efficiency. We chose to keep the convolution layers fixed to retain their feature extraction abilities, from the ImageNet dataset using a selective layer freezing technique. For the four layers of the model, we enabled tuning to customize it for our classification task and make use of the existing pre-trained weights. To maintain training stability. Ensure progress in learning we included batch normalization, across the entire network. Furthermore, we included a system for adjusting the learning rate as needed during training to ensure the convergence without getting stuck in local minimum points.

Texture Feature Extraction via GLCM

Enhancing the characteristics further involves utilizing Gray Level Cooccurrence Matrix (GLCM) which helps in capturing detailed texture designs through analysis methods we apply in our implementation of GLCM.

- Analyzing parameters: Distance represented by vectors is measured in pixels, as [1,2,3,4].
- Angular positions include 4 orientations, at 45 degrees intervals starting from 1, to 3 o'clock.
- Calculating Features: Contrast: Differences, in measuring the changes in intensity, at locations.
- Correlation: Exploring the connection between pixels through the evaluation of relationships.
- Energy: The concept of energy is a way to measure how consistent something is, in terms of its texture.
- Homogeneity: Assessing uniformity levels, in the texture of an area.

This setup results in 16 texture characteristics, for each image. Offers detailed spatial information, on various scales and orientations.

Feature Fusion and Classification

At the end of our pipeline is a system for integrating features and classifying them effectively. During the feature integration phase, we merge flattened VGG16 features with GLCM features by concatenating them using a fusion strategy. To maintain stability in computations and avoid any feature overpowering the classification process we include a step of normalizing the features. The resulting unified feature vectors capture both learning based features and traditional textural qualities to represent the image content, in detail. For the task of classification, we use the CatBoost algorithm with tuned hyperparameters that were found through thorough grid search optimization efforts. The setup consists of 907 iterations, a learning rate of 0.01837, a tree depth of 5 and an L2 leaf regularization value set at 1.57 e^07. We opt for the MVS (Multivariate Split) bootstrap method, with a strength set at 7.46 and a bagging temperature of 6.61 e^07. To avoid overfitting we have included a stopping strategy using 'Iter' type, with a waiting period of 22 iterations. Through validation, across parameters and meticulous fine-tuning procedures, in place to prevent overfitting concerns effectively whilst enhancing classification accuracy reliably; this leads to a model that can generalize adeptly to unfamiliar data instances.

Results

The CatBoost model stands out with an accuracy of 0.830 with high precision, recall with a balance in F1 score. We also found out that complex dependencies and patterns were not captured by KNN and AdaBoost, but the conventional SVM and Random Forest gave decent results. A shift towards hybrid models began with CNN + CatBoost that initially resulted in poor performance due to suboptimal feature extraction. Notably, we got a jump to 0.870 by using Gray-Level Co-Occurrence Matrix (GLCM) features plugged with CNN + CatBoost + GLCM model. This result emphasizes the need of feature engineering to improve the custom tuned hybrid models in complex classification tasks.

A class-wise sensitivity analysis was carried out to check whether the model can distinguish between different classes with the help of confusion matrix. We observed lower recall values in some classes although they had high sensitivity so there was a need to enhance performance of some classes by fine-tuning. This observation was visible from per class-recall evaluation.

Here, CatBoost maintained a trade-off between precision and recall that outperformed other models. CatBoost avoided overfitting to any one category and predicted correct classes which is evident from its balanced metrics. Overall CatBoost performs consistently capable of processing optimal features making it well suited to handle dataset's complexity. To improve recall for individual classes we could create ensemble models or adjust hyperparameters.

Discussion

During the research project conducted here different machine learning methods were assessed based on their performance, in classification tasks. Among these models examined, CatBoost stood out as the leading contender delivering results with an accuracy rate of 83% coupled with precision recall rates and F1 scores. The comparison of performance of classifiers is depicted in Table II.

TABLE II : comparison of classifiers

MODEL	ACC.	PRECISION	F1 SCORE	RECALL
SVM	0.824	0.824	0.823	0.823
K-NN	0.644	0.650	0.643	0.647
DECISION TREE	0.816	0.817	0.815	0.815
RANDOM FOREST	0.820	0.820	0.819	0.819
GRADIENT BOOSTING	0.788	0.788	0.788	0.789
ADBOOST	0.706	0.706	0.706	0.705
NAIVE BAYES	0.756	0.758	0.756	0.757
CATBOOST	0.830	0.830	0.829	0.829
CNN+RANDOM FOREST	0.804	0.805	0.803	0.802
CNN CATBOOST	0.487	0.461	0.427	0.475

CNN GLCM +CATBOOST	**0.870**	**0.873**	**0.869**	**0.868**

Additionally, conventional models such as SVM and Random Forest also showcased performances with accuracies of 82% and 82% respectively. In contrast the k nearest neighbors (k NN) model struggled in achieving accuracy 64% due to its difficulties in handling decision boundaries. Results from the analysis showed that Naive Bayes had an accuracy of 75% indicating a balance between precision and recall compared to AdaBoost with 70% accuracy in the study of Gradient Boosting models and Decision Trees effectiveness. The models that combined CNN with machine learning techniques had success rates; for example, CNN combined with Random Forest achieved a respectable 80% accuracy while CNN paired with CatBoost struggled at only 48% possibly due to difficulties in feature extraction. The performance saw an improvement, with the integration of texture features using Gray Level Co-Occurrence Matrix (GLCM) in the model involving CNN + CatBoost + GLCM which resulted in the overall accuracy of (0.870). This highlights the significance of incorporating domain features to enhance classification accuracy The combined approach of CatBoost with CNN and GLCM outperformed other models indicating that feature engineering plays a crucial role in boosting model performance.

Conclusion

The accurate detection of fetal brain abnormalities in low-quality ultrasound images remains a critical challenge, particularly in resource-constrained healthcare environments where access to high-end imaging tools and radiological expertise is limited. To address this gap, this study proposed a hybrid framework integrating Convolutional Neural Networks (CNN), Gray-Level Co-occurrence Matrix (GLCM) texture analysis, and CatBoost classification, specifically designed to enhance diagnostic accuracy in low-quality prenatal ultrasound images. Our methodology introduced bilateral filtering for preprocessing to mitigate noise while preserving anatomical edges, a modified VGG16 architecture for deep feature extraction, and GLCM-based texture descriptors to capture spatial patterns imperceptible to CNNs alone. By fusing these features and leveraging CatBoost's robust handling of categorical data and gradient boosting, the model achieved an accuracy of 87%, with precision, recall, and F1-score all exceeding 86%. These results demonstrate a significant improvement over baseline models such as standalone CNNs (48% accuracy) and traditional machine learning classifiers (e.g., SVM at 82%), validating the necessity of combining handcrafted texture features with deep learning for low-quality image analysis. Practically, this approach can empower healthcare providers in rural or under-resourced regions by offering a reliable, automated tool for early anomaly detection, thereby reducing dependency on specialized radiologists and improving perinatal care accessibility.

Limitations and Future Work

While the proposed hybrid CNN-GLCM framework demonstrates promising performance in detecting fetal brain abnormalities in low-quality images, several limitations must be acknowledged. First, the dataset was limited to trans-cerebellar views due to quality inconsistencies in other anatomical planes, potentially restricting the model's generalizability to multi-view clinical scenarios. Second, class-specific recall imbalances were observed, indicating that rare anomalies might be under-detected. Finally, the framework's reliance on handcrafted texture descriptors may require recalibration for diverse noise patterns encountered across varying ultrasound devices.

Future research should prioritize validating the framework on multi-view datasets, such as the IIT Madras DHARANI dataset, which includes diverse fetal brain regions. Extending the model to 3D ultrasound analysis and incorporating ensemble architectures could address class-specific performance gaps. Techniques like adaptive data augmentation or cost-sensitive learning may further mitigate recall imbalances. Investigating lightweight neural networks or edge-computing optimizations could enhance deployability in low-resource clinics. Lastly, exploring explainable AI components would improve clinician trust and facilitate integration into prenatal diagnostic workflows.

References

1. C. Palma dos Reis, S. Brás, T. Meneses, A.S. Cerdeira, M. Vatish, and A.T. Martins, "SFlt1/PlGF ratio as predictor of fetal deterioration in early growth restriction," in Int. Conf. Obstet. Gynecol., 2023, pp. 635-643, doi: 10.1111/aogs.14546.
2. X. Wang, W. Wang, and X. Cai, "Automatic measurement of fetal head circumference using a novel GCN-assisted deep convolutional network," in Int. Conf. Comput. Biol. Med., 2022, pp. 105515, doi: 10.1016/j.compbiomed.2022.105515.
3. Z. Lin, S. Li, D. Ni, Y. Liao, H. Wen, J. Du, S. Chen, T. Wang, and B. Lei, "Multi-task learning for quality assessment of fetal head ultrasound images," in Int. Conf. Med. Image Anal., 2019, pp. 101548, doi: 10.1016/j.media.2019.101548.
4. K. Arun, S. Phaneesh, S. Reddy, S. Sreeman, and Y. Ghildiyal, "An information system on fetal health classification based on CNN and hybrid-CNN with dimensionality reduction," in E3S Web Conf., 2023, pp. 01029, doi: 10.1051/e3sconf/202343001029.

5. R. Arnaout, L. Curran, Y. Zhao, J.C. Levine, E. Chinn, and A.J. Moon-Grady, "Neural network ensemble for expert-level detection of congenital heart disease," in Nat. Med. Conf., 2021, pp. 882-891, doi: 10.1038/s41591-021-01342-5.
6. M.A. Duhayyim, S. Abbas, A.A. Hejaili, N. Kryvinska, A. Almadhor, and H. Mughal, "Ensemble learning for fetal health classification," in Comput. Syst. Sci. Eng. Conf., 2023, pp. 823-842, doi: 10.32604/csse.2023.037488.
7. H.N. Xie, N. Wang, M. He, L.H. Zhang, H.M. Cai, J.B. Xian, M.F. Lin, J. Zheng, and Y.Z. Yang, "Deep-learning algorithms for fetal brain ultrasound classification," in Int. Conf. Ultrasound Obstet. Gynecol., 2020, pp. 579-587, doi: 10.1002/uog.21967.
8. J. Tang, J. Han, B. Xie, J. Xue, H. Zhou, Y. Jiang, L. Hu, C. Chen, K. Zhang, F. Zhu, and L. Ju, "Two-stage ensemble learning model for screening fetal genetic diseases," in Int. Conf. Environ. Res. Public Health, 2023, pp. 2377, doi: 10.3390/ijerph20032377.
9. P.A. Baird, T.W. Anderson, H.B. Newcombe, and R.B. Lowry, "Population study of genetic disorders in children and young adults," in Am. Conf. Human Genetics, 1988, pp. 677-693.
10. X. Liu, Z. Dong, H. Li, J. Ren, H. Zhao, H. Li, W. Chen, and Z. Xiao, "H-RNet: Hybrid relation network for few-shot learning-based classification," in Remote Sens. Conf., 2023, pp. 2497, doi: 10.3390/rs15102497..
11. M.C. Walker, I. Willner, O.X. Miguel, M.S.Q. Murphy, D. El-Chaâr, and F. Moretti, "Deep-learning in fetal ultrasound analysis for first trimester cystic hygroma diagnosis," in Int. Conf. Med. Imaging, 2022, pp. e0269323, doi: 10.1371/journal.pone.0269323.
12. V. Singh, P. Sridar, J. Kim, R. Nanan, N. Poornima, S. Priya, G. Sameera, S. Chandrasekaran, and R. Krishnakumar, "CNN-based semantic segmentation of cerebellum in fetal ultrasound images," in IEEE Access Conf., 2021, pp. 1-1, doi: 10.1109/access.2021.3088946.
13. X.P. Burgos-Artizzu et al., "Deep convolutional neural networks for maternal fetal ultrasound plane classification," in Sci. Rep. Conf., 2020, pp. 10200, doi: 10.1038/s41598-020-67076-5.