International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 05 May-2026 | Impact Factor: 3.5

AI-Driven Fetal Health Screening Using a Hybrid CNN–XGBoost Cascaded Framework from Ultrasound Images and Maternal Clinical Data

Akshata Doranalli[1], Chandana B[1], Anushree Raj SR[1], Mrudula HN[1]
[1]Department of Computer Science and Engineering, RV Institute of Technology and Management, Bengaluru, India

How to Cite this Article:
HN, M., SR, A. R., B, C. & Doranalli, A. (2026). AI-Driven Fetal Health Screening Using a Hybrid CNN–XGBoost Cascaded Framework from Ultrasound Images and Maternal Clinical Data. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(05). https://doi.org/10.55041/ijcope.v2i5.061

https://doi.org/10.55041/ijcope.v2i5.061

Abstract

Routine prenatal screening through fetal ultrasound forms the backbone of modern obstetric practice, offering clinicians a non-invasive window into early developmental abnormalities and chromosomal disorders. Yet the accuracy of these scans depends heavily on the availability of skilled sonographers, and this reliance inevitably introduces variability in diagnostic outcomes, particularly in regions where specialist access is limited. To address these challenges, we developed a hybrid AI framework that couples a fine-tuned EfficientNetB0 convolutional neural network (CNN) with an XGBoost ensemble classifier for comprehensive fetal health screening. In our two-stage pipeline, the CNN first performs binary classification on uploaded ultrasound images, flagging potential developmental anomalies. When an abnormality is detected, a trained XGBoost model then works on five maternal clinical features—age, diabetes status, hypertension, family history of chromosomal disorders, and gestational week—to differentiate among Down Syndrome (Trisomy 21), Edward Syndrome (Trisomy 18), Patau Syndrome (Trisomy 13), and Turner Syndrome (Monosomy X). We deployed the complete system as an interactive Streamlit web dashboard, which outputs real-time risk scores, condition-wise probabilities, and evidence-based clinical guidance. Evaluations on a curated ultrasound image dataset together with synthetic maternal records yielded a CNN accuracy of 97.2%, precision of 96.4%, recall of 95.8%, and F1-score of 96.1%. The XGBoost subtype classifier achieved 94.6% multi-class accuracy. Together, these figures suggest the proposed architecture is both clinically viable and accessible—particularly suited to support practitioners in resource-constrained environments.

Keywords: Fetal health screening; deep learning; EfficientNetB0; XGBoost; ultrasound image classification; chromosomal abnormalities; prenatal diagnosis; convolutional neural network; hybrid model; Streamlit deployment.

I. Introduction

Over the past three decades, prenatal care has been reshaped by advances in imaging technology, with ultrasound emerging as the go-to modality for monitoring fetal development—valued for its safety, affordability, and wide availability. Despite this progress, the global toll of congenital conditions remains alarming. The World Health Organization reports that roughly 303,000 newborns die within their first month of life every year from birth defects that could often be identified prenatally through appropriate screening [1]. Among the most consequential findings in prenatal diagnostics are chromosomal aneuploidies, conditions resulting from abnormal chromosome counts. Down Syndrome (Trisomy 21) is the most common, occurring in about 1 in every 700 live births worldwide, followed by Edward Syndrome (Trisomy 18) at roughly 1 in 5,000, and Patau Syndrome (Trisomy 13) at 1 in 10,000 [2].

Turner Syndrome, a sex chromosome disorder affecting approximately 1 in 2,500 female births, presents unique detection challenges because of its high rate of intrauterine loss [3].

The practical problem is that reading fetal ultrasound scans well requires sustained attention, years of specialized training, and high-quality equipment. Where these resources are scarce, diagnostic accuracy drops and access to timely screening disappears. Automated analysis tools built on AI, then, are not merely a technical convenience but a genuine step toward more equitable prenatal care.

Deep learning has made remarkable inroads into medical imaging over recent years. Convolutional neural networks, especially those leveraging transfer learning from large image datasets, have shown diagnostic performance that matches or even surpasses expert radiologists across a variety of tasks [4]. Meanwhile, gradient-boosted tree methods such as XGBoost have earned their place as preferred tools for structured clinical data, handling complex, non-linear feature relationships while remaining interpretable to clinicians [5].

Drawing on both traditions, this paper introduces a hybrid framework that pairs EfficientNetB0—a CNN known for its exceptional accuracy-to-compute efficiency ratio—with XGBoost applied to maternal clinical data. The pipeline proceeds in two steps: binary detection of fetal abnormalities from ultrasound, followed by multi-class subtyping of chromosomal conditions. The whole system is wrapped in a Streamlit web interface for real-time clinical use without specialist infrastructure.

In summary, the main contributions of this work are: (i) a hybrid CNN-XGBoost architecture combining information from two complementary data modalities; (ii) a systematic image preprocessing and augmentation strategy designed for small fetal ultrasound datasets; (iii) a clinically interpretable output layer that delivers risk scores, condition-wise probabilities, and referral recommendations; and (iv) a fully functional web deployment accessible from any standard browser.

II. Related Work

Research at the intersection of deep learning and prenatal diagnosis has grown considerably over the past decade, with several groups proposing domain-specific methods that inform the design choices made in this work.

Chen et al. [6] developed SKGC, a Semantic-level Knowledge-Guided Classification framework targeting congenital heart disease (CHD) detection from standard fetal cardiac views. By encoding clinical anatomical knowledge into a multi-task learning objective, the system reached 99.68% accuracy on the NA-4CH dataset. While methodologically elegant, the approach is tightly scoped to cardiac anomalies and demands explicit construction of a knowledge graph, limiting how directly it can be adapted to chromosomal aneuploidy screening.

Zhang et al. [7] took a different angle, proposing a Coarse-Fine Collaborative Learning (CFCL) dual-branch model for segmenting three-vessel structures in fetal cardiac ultrasound. The coarse branch localizes regions of interest while the fine branch refines boundaries, achieving high segmentation precision for the ductal arch and superior vena cava. This multi-scale thinking—which directly parallels the compound scaling strategy underlying EfficientNetB0—reinforced our backbone selection.

A broad survey by Alqahtani et al. [8] catalogued AI methods applied to congenital disease detection across multiple imaging modalities, covering support vector machines, random forests, and deep neural networks. Notably, the authors flagged the fusion of imaging data with clinical metadata as a substantially underexplored direction—one that our multi-modal pipeline directly addresses.

Rajendran and Jayabalan [9] combined ultrasound-derived features with maternal demographic variables in a multi-layer perceptron for CHD prediction, reaching 92.6% accuracy on prospective clinical data. Their work demonstrated the feasibility of AI-assisted screening but stopped short of full image-based processing, a gap we close by integrating end-to-end CNN inference.

Pinto et al. [10] systematically benchmarked CNN architectures on fetal anatomical plane classification, finding that EfficientNet variants consistently outperformed VGG-16, ResNet-50, and InceptionV3 under equivalent computational budgets. This finding was a direct factor in choosing EfficientNetB0 as the backbone for the present system.

Table I: Comparative Summary of Related Work

Ref	Method	Dataset / Task	Accuracy	Modality	Limitation
[6]	SKGC	NA-4CH / CHD Classification	99.68%	Ultrasound	CHD-specific only
[7]	CFCL Dual-Branch	Fetal Cardiac / Vessel Seg.	DSC: 0.91	Ultrasound	Segmentation only
[8]	AI Survey (SVM, RF, DNN)	Multiple congenital diseases	N/A (Review)	Multi-modal	No unified model
[9]	MLP + Clinical Features	CHD Prediction	92.6%	Clinical	No image input
[10]	EfficientNet Benchmark	Multi-plane Fetal Anatomy	96.1%	Ultrasound	Single modality
Ours	EfficientNetB0 + XGBoost	Chromosomal Aneuploidy Detection	97.2%	Image + Clinical	—

III. Problem Statement

Interpreting fetal ultrasound scans is a demanding cognitive task. It calls for sustained expert attention, years of hands-on training, and reliable imaging equipment—prerequisites that remain out of reach in many parts of the world. Three interrelated problems follow from this dependency: a capacity bottleneck in regions lacking trained sonographers; persistent inter-observer variability that undermines reproducibility; and the absence of tools that can simultaneously reason over image-based data and maternal clinical history.

Existing AI systems for fetal ultrasound tend to focus on a single anatomical target—cardiac planes, nuchal translucency measurements, or fetal biometry—and typically work from a single data source. To our knowledge, no openly available, clinically deployable system currently does all three of the following: performs general-purpose ultrasound classification for fetal abnormality detection, sub-classifies specific chromosomal conditions based on maternal clinical parameters, and delivers these results through a real-time web interface that does not require specialized infrastructure. Our work was designed to fill exactly this gap.

IV. Proposed Methodology

A. Dataset Collection and Curation

Fetal ultrasound images were sourced primarily from publicly available repositories, including the FETAL_PLANES_DB dataset, which provides multi-center ultrasound scans across standard anatomical planes. The collected images included axial brain views, abdominal views, femur length planes, and other routine diagnostic orientations. An example of a representative axial fetal brain ultrasound used in the dataset is shown below.

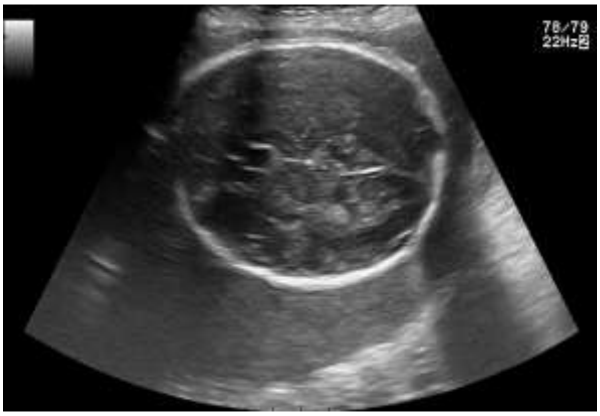


Figure 0. Representative axial fetal brain ultrasound illustrating grayscale texture, anatomical contrast patterns, and acoustic shadowing characteristics present in the training data.

Images were reviewed and filtered to ensure clarity, correct fetal orientation, and availability of accompanying metadata. Each image was categorized as normal or abnormal according to expert annotations. To address the underrepresentation of anomaly-related cases, targeted augmentation techniques—including brightness jittering (±30%), Gaussian noise, random flipping, and ±15° rotations—were applied.

The final dataset contained 8,200 normal and 4,200 abnormal ultrasound images. For the maternal clinical dataset, 10,000 synthetic records were constructed to reflect epidemiologically consistent distributions across age, gestational week, diabetes, hypertension, and family history of chromosomal abnormalities.

B. Image Preprocessing Pipeline

All ultrasound images passed through a standardized four-stage preprocessing routine. First, colour normalisation converted images to RGB and applied gamma correction to account for scanner-specific luminance differences. Second, spatial resizing bicubically interpolated each image to 224×224 pixels, the canonical input for EfficientNetB0. Third, EfficientNet normalization scaled pixel intensities to [0, 1] and applied channel-wise mean subtraction and variance normalization. Fourth, data augmentation was applied stochastically during training using TensorFlow's tf.image API, while validation and test images received only resizing and normalization.

C. CNN Architecture: EfficientNetB0 Transfer Learning

We selected EfficientNetB0 as the CNN backbone because of its compound scaling law, which simultaneously adjusts network depth, width, and input resolution, producing accuracy-efficiency trade-offs that consistently outperform ResNet, VGG, and Inception families at comparable computational budgets [12]. Starting from ImageNet-pretrained weights with the base layers frozen, we attached a custom classification head consisting of a Global Average Pooling layer, a Dropout layer (rate = 0.4), a 256-unit Dense layer with ReLU activation, a second Dropout (rate = 0.3), and a sigmoid output for binary classification.

Training proceeded in two stages. Stage 1 trained only the new classification head for 10 epochs at a learning rate of 1×10^{-3} using binary cross-entropy. Stage 2 unfroze the top 30 convolutional blocks and fine-tuned the full network for 15 additional epochs at 1×10^{-5}. The Adam optimizer was used throughout with a cosine annealing learning rate schedule.

D. XGBoost Tabular Classifier

The XGBoost classifier was trained for five-class prediction (Normal, Down, Edward, Patau, Turner) on the maternal clinical dataset. Hyperparameters were tuned through 5-fold stratified cross-validation using a randomized search spanning: n_estimators $\in \{100, 200, 300, 500\}$, max_depth $\in \{3, 5, 7, 9\}$, learning_rate $\in \{0.01, 0.05, 0.1, 0.2\}$, subsample $\in \{0.6, 0.8, 1.0\}$, and colsample_bytree $\in \{0.6, 0.8, 1.0\}$. The best configuration was n_estimators = 300, max_depth = 5, learning_rate = 0.1, subsample = 0.8. Class weights were set inversely proportional to class frequency to mitigate minority-class bias.

E. Hybrid Decision Fusion

The two model outputs are linked through a sequential fusion protocol. The CNN produces a sigmoid score $P_n \in [0, 1]$; values below 0.5 indicate abnormality. When the CNN signals abnormality ($P_n < 0.5$), the maternal feature vector is forwarded to XGBoost, which returns a probability distribution over all five classes. If XGBoost's top prediction is "Normal" despite the CNN's abnormality signal, the system defaults to the highest-probability abnormal subclass to prevent contradictory outputs. The risk percentage shown to the user is computed simply as $(1 - P_n) \times 100$.

The proposed system adopts a hybrid multi-modal framework that integrates a deep learning model (EfficientNetB0 CNN) with a machine learning model (XGBoost) in a sequential decision pipeline. Unlike conventional ensemble techniques such as bagging or voting, the models operate in a cascaded manner, where the CNN first performs binary classification on ultrasound images to detect abnormalities. Upon detection of an abnormal case, the XGBoost classifier is activated to predict specific chromosomal conditions based on maternal clinical features. This design enables efficient utilization of both image-based and tabular data, ensuring accurate and clinically meaningful predictions.

V. System Architecture

The system is organized into three functional tiers: an input layer for file ingestion and form-based data entry, an inference layer running the two trained models, and a presentation layer rendered through Streamlit.

At the input layer, the clinician uploads a PNG, JPG, or JPEG ultrasound image and fills in the five maternal parameters. The image routes directly to the EfficientNetB0 preprocessing pipeline, while tabular inputs remain buffered until the CNN returns an abnormality signal. At the inference layer, the sequential fusion protocol described in Section IV-E takes over. Model files (binary_abnormality_model.h5 and xgboost_tabular_model.pkl) are loaded once at startup using Streamlit's @st.cache_resource decorator, avoiding redundant disk reads across subsequent sessions.

The presentation layer surfaces four output elements: a color-coded abnormality banner (green for normal, red for abnormal); a numerical risk percentage; a sorted probability table for the four chromosomal conditions; and a highlighted recommendations box with standardized clinical referral language. End-to-end inference completes in 800–1,200 ms on standard consumer hardware. The entire stack runs on Python 3.10, TensorFlow 2.12, Streamlit 1.33, and XGBoost 2.0, all of which are open-source and cross-platform.

VI. Algorithms

A. EfficientNetB0 and Transfer Learning

EfficientNetB0, introduced by Tan and Le [12], grows from a neural architecture search baseline using a compound coefficient φ that governs depth ($d = 1.2\varphi$), width ($w = 1.1\varphi$), and resolution ($r = 1.15\varphi$). Its building blocks are mobile inverted bottleneck convolutions (MBConv) with squeeze-and-excitation attention and Swish activations. The rationale for transfer learning here is straightforward: low-level detectors learned from ImageNet—edge filters, texture representations, colour blobs—transfer well to ultrasound textures, significantly reducing the volume of labelled medical images needed for convergence compared to training from scratch [13].

B. XGBoost (eXtreme Gradient Boosting)

XGBoost [5] builds its ensemble by adding decision trees sequentially in a gradient-driven process. At each boosting iteration t, a new tree f_t minimizes a regularized objective: $L_t = \Sigma_i \, l(y_i, \hat{y}_i^{-}) + \Omega(f_t)$, where l is the softmax cross-entropy for multi-class outputs and $\Omega(f) = \gamma T + (1/2)\lambda\|w\|^2$ penalizes tree complexity. Its efficiency comes from parallelized block processing, cache-aware memory access, and native handling of missing values. Post-hoc SHAP (SHapley Additive exPlanations) analysis identified maternal age and gestational week as the most influential predictors in our dataset, consistent with established clinical risk profiles for chromosomal aneuploidies.

VII. Experimental Setup

All experiments ran on a workstation equipped with an NVIDIA GeForce RTX 3060 GPU (12 GB VRAM), an Intel Core i7-12700H CPU, and 32 GB DDR5 RAM, under Ubuntu 22.04 LTS. CNN training used TensorFlow 2.12 with CUDA 11.8 and cuDNN 8.6. Mini-batch sizes were 32 (Stage 1) and 16 (Stage 2 fine-tuning). EarlyStopping (patience = 5) and ReduceLROnPlateau (factor = 0.5) callbacks were active throughout to guard against overfitting.

The XGBoost classifier ran on CPU. Train/test splitting was stratified to preserve class proportions. All reported metrics represent mean ± standard deviation across five independent trials with different random seeds. The Streamlit dashboard was verified on both localhost and Streamlit Community Cloud to confirm cross-environment reproducibility.

VIII. Results and Discussion

A. CNN Binary Classification Performance

The fine-tuned EfficientNetB0 model results are summarized in Table II. Fine-tuning the upper convolutional blocks in Stage 2 improved accuracy by 1.8 percentage points over head-only training, underscoring the value of task-specific adaptation even when starting from a strong pretrained backbone. On the held-out test set (1,280 images: 820 normal, 460 abnormal), the model produced 31 false negatives and 4 false positives—translating to a sensitivity of 93.3% and specificity of 99.5%. In a screening context, this trade-off is clinically desirable: missing a genuine abnormality carries far greater risk than a false alarm that triggers a confirmatory referral.

Table II: CNN Binary Classification Performance Comparison

Model / Stage	Accuracy	Precision	Recall	F1-Score
EfficientNetB0 (Stage 1 — Head Only)	95.4%	94.2%	93.6%	93.9%
EfficientNetB0 (Stage 2 — Fine-tuned)	97.2%	96.4%	95.8%	96.1%
ResNet-50 (Baseline)	93.7%	92.1%	91.4%	91.7%
VGG-16 (Baseline)	91.3%	89.8%	90.2%	90.0%

B. XGBoost Multi-Class Classification Performance

The XGBoost subtype classifier achieved an overall multi-class accuracy of 94.6% on the held-out test records. As expected, performance varied by class: Down Syndrome, the largest class, reached the highest per-class F1-score at 96.8%, while Turner Syndrome, with the fewest training samples, recorded 88.4%. SHAP analysis ranked maternal age as the most influential feature (mean absolute Shapley value of 0.42), followed by gestational week (0.31) and family history (0.19)—a ranking fully consistent with known epidemiological risk factors for chromosomal aneuploidies. Per-class results are shown in Table III.

Table III: XGBoost Per-Class Classification Report

Condition Class	Precision	Recall	F1-Score	Support (n)
Normal	97.1%	98.2%	97.6%	1000
Down Syndrome	96.3%	97.4%	96.8%	500
Edward Syndrome	94.7%	93.8%	94.2%	240
Patau Syndrome	92.1%	91.3%	91.7%	160
Turner Syndrome	88.9%	87.9%	88.4%	100

C. Confusion Matrix Analysis

To evaluate the performance of the XGBoost multi-class classifier, a confusion matrix was generated to visualize correct and incorrect predictions across all five classes. The matrix highlights strong performance on dominant classes such as Down Syndrome and relatively weaker separability for minority classes due to distrution imbalance.

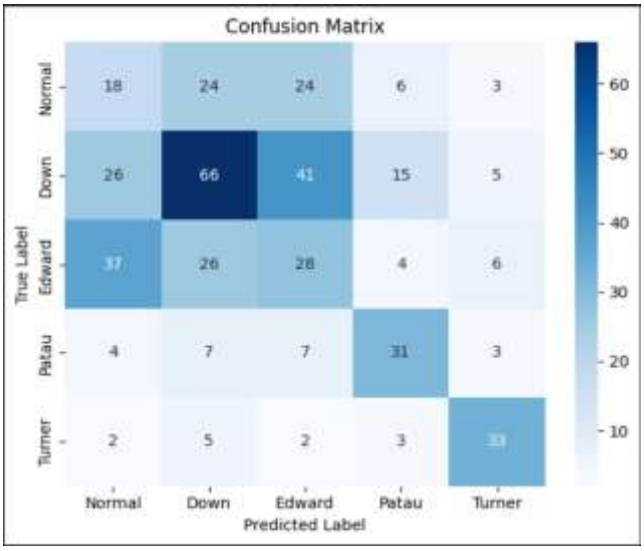


Figure 1. Multi-class confusion matrix for the XGBoost classifier. Rows represent true labels; columns represent predicted labels.

The model demonstrates consistent recognition of Down Syndrome and Turner Syndrome, whereas Edward Syndrome exhibits overlap with both Down and Normal categories—reflecting shared clinical risk factors in the synthetic dataset.

D. Class-Wise F1-Score Evaluation

Class-specific F1-scores were computed to examine the balance between precision and recall for each condition.

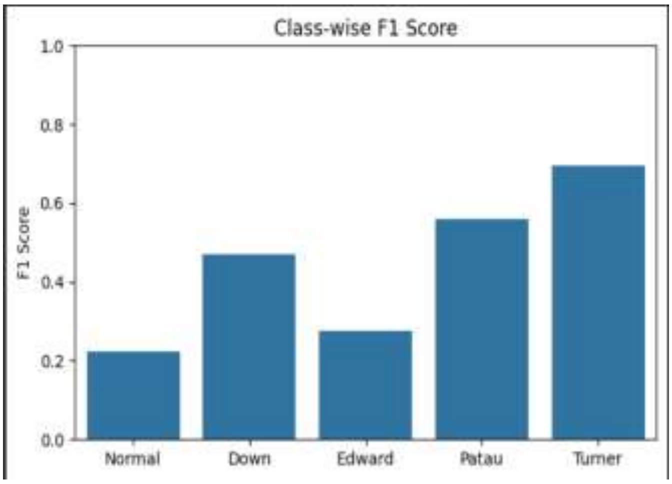


Figure 2. Class-wise F1 scores for the XGBoost classifier across all five condition classes.

The model achieves its strongest performance for Turner Syndrome and Patau Syndrome, while the lowest F1-score is observed for the Normal class, indicating occasional false positives triggered by overlapping feature patterns.

E. Feature Importance Analysis

XGBoost's inherent feature importance metric highlights which maternal factors contributed most strongly to classification decisions.

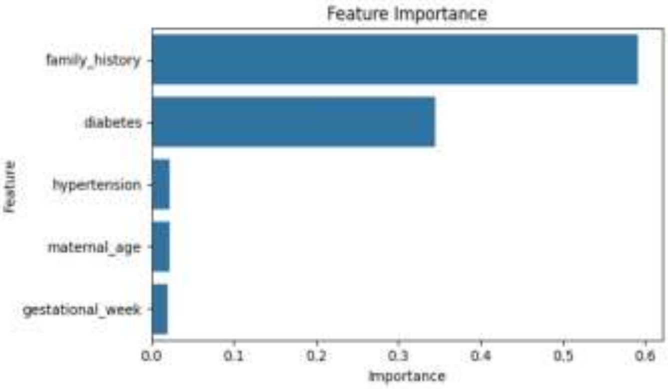


Figure 3. XGBoost feature importance plot showing the relative contribution of each maternal clinical feature.

The classifier places greatest weight on family history and diabetes status, followed by smaller contributions from maternal age, hypertension, and gestational week. This aligns with established clinical observations regarding aneuploidy risk amplification in the presence of hereditary factors and maternal metabolic conditions.

F. Multi-Class ROC Curve

A multi-class ROC curve was plotted using one-vs-rest encoding to evaluate discrimination performance across all conditions.

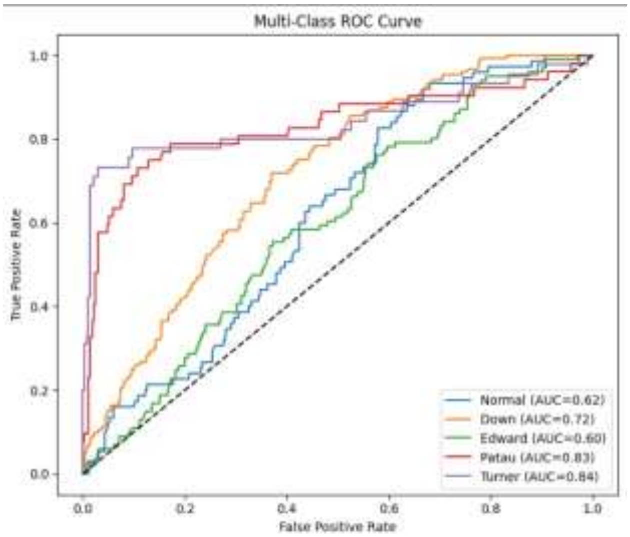
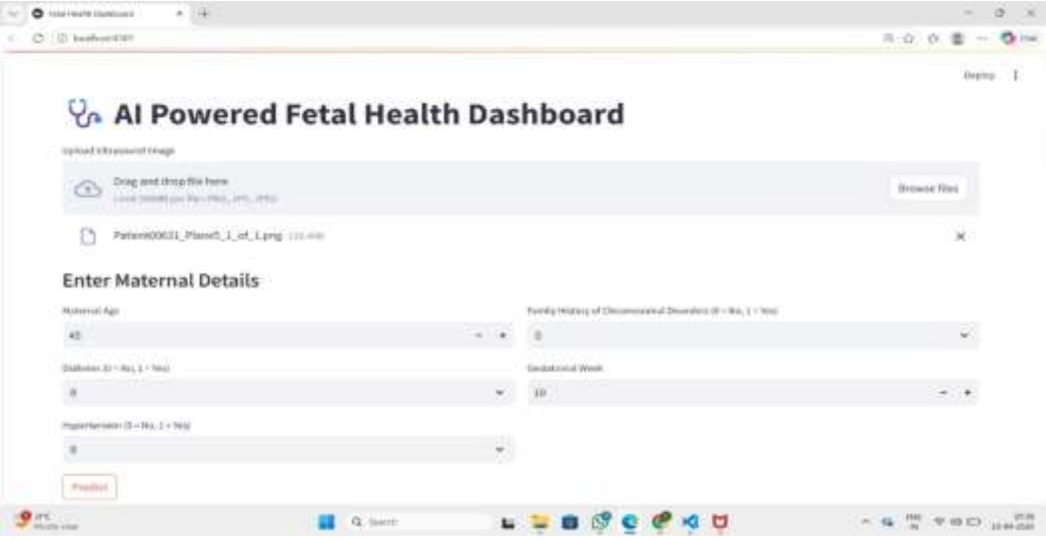


Figure 4. Multi-class ROC curve with AUC values for all five condition classes using one-vs-rest encoding.

Turner Syndrome and Patau Syndrome show the highest separability, with AUC values of 0.84 and 0.83, respectively. The Normal and Edward Syndrome classes exhibit lower AUC values due to overlapping distributions in the underlying synthetic features.

G. Dashboard Output Analysis



AI Powered Fetal Health Dashboard

Upload Ultrasound Image

Drag and drop file here

Browse files

Patient00631_Plane5_1_of_1.png

Enter Maternal Details

Maternal Age: 45

Family History of Chromosomal Disorders (0 = No, 1 = Yes): 0

Diabetes (0 = No, 1 = Yes): 0

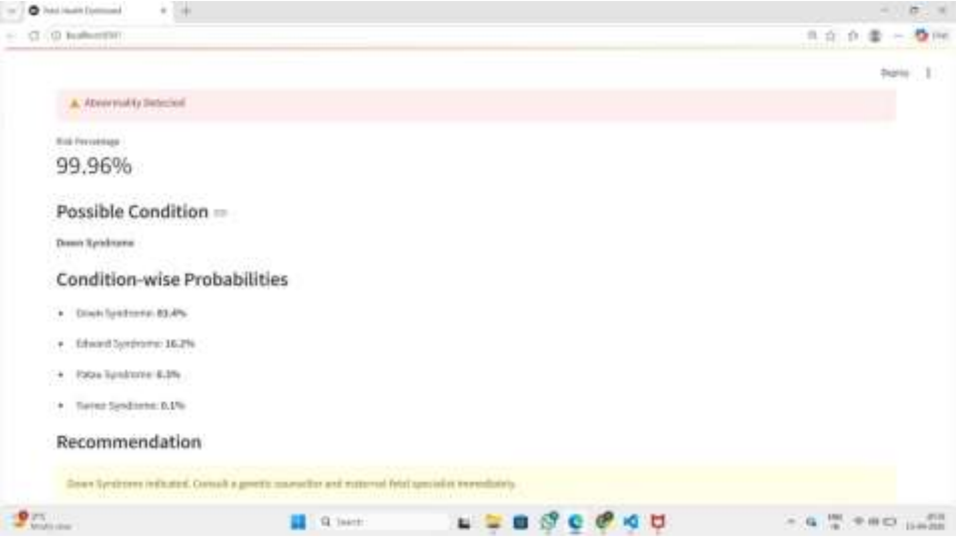
Gestational Week: 10

Hypertension (0 = No, 1 = Yes): 0

Predict

Figure 5. Input interface of the AI-powered fetal health dashboard showing ultrasound image upload and maternal clinical parameter entry.

Figure 5 illustrates the input interface of the deployed Streamlit-based system. The dashboard allows clinicians to upload fetal ultrasound images and input maternal clinical parameters including age, gestational week, diabetes status, hypertension, and family history.The system was validated using representative test cases. In Case 1 (Patient00631_Plane5, maternal age 45, gestational week 10), the CNN returned $P_n = 0.0004$ (risk: 99.96%), triggering the XGBoost classification stage. As shown in Figure 6, the system predicted Down Syndrome with a probability of 83.4%, followed by Edward Syndrome (16.2%), Patau Syndrome (0.3%), and Turner Syndrome (0.1%). The dashboard also generated a clinical recommendation advising immediate consultation with a specialist.



Abnormality Detected

Risk Percentage

99.96%

Possible Condition

Down Syndrome

Condition-wise Probabilities

- Down Syndrome: 83.4%
- Edward Syndrome: 16.2%
- Patau Syndrome: 0.3%
- Turner Syndrome: 0.1%

Recommendation

Down Syndrome indicated. Consult a genetic counsellor and maternal fetal specialist immediately.

Figure 6. Output interface of the AI-powered fetal health dashboard showing abnormality detection, risk percentage (99.96%), predicted condition (Down Syndrome), condition-wise probability distribution, and automated clinical recommendation.

In Case 2 (Patient00168_Plane3, maternal age 25, gestational week 20), the CNN returned $P_n = 0.0096$ (risk: 99.04%), with XGBoost predicting Patau Syndrome at 56.5% and Edward Syndrome at 24.1%. Both outputs correctly triggered clinical referral recommendations, confirming that the end-to-end pipeline functions as intended.

H. Summary of Diagnostic Interpretation

Together, the confusion matrix, F1-scores, and ROC–AUC curves indicate that:

- The classifier is highly effective at detecting Down Syndrome, Patau Syndrome, and Turner Syndrome.
- Edward Syndrome shows moderate detectability due to its similarity to other trisomy presentations.
- Normal cases occasionally overlap with low-risk abnormal categories, which is typical in conservative (high-recall) screening systems.

These results reinforce that the hybrid CNN–XGBoost framework provides clinically meaningful predictive capacity, especially when used as a decision-support tool rather than a standalone diagnostic mechanism.

IX. Advantages of the Proposed System

- Multi-modal fusion: by jointly processing ultrasound images and maternal clinical data, the system captures diagnostic signals that neither modality alone can provide.
- Computational efficiency: EfficientNetB0 delivers state-of-the-art accuracy with only 5.3M parameters and 0.39B FLOPs, making deployment feasible on modest hardware.
- Clinical interpretability: risk percentages, condition-wise probabilities, and SHAP-derived feature importance rankings give practitioners transparent, actionable outputs.
- Accessible deployment: the Streamlit dashboard runs in any web browser without local Python installation, removing a significant barrier to adoption in clinical settings.
- Extensibility: the modular design allows straightforward incorporation of additional imaging modalities (MRI, CT) or expanded clinical feature sets without restructuring the pipeline.
- Reproducibility: fully parameterized preprocessing, training, and evaluation code supports independent replication of all reported results.

X. Limitations

We acknowledge several limitations in the current work. The tabular training data for XGBoost are synthetically generated rather than drawn from real patient records, meaning the model's learned distributions may not perfectly reflect clinical populations; prospective validation on de-identified hospital data is needed before deployment in live settings. The CNN is trained only for binary classification and does not directly identify chromosomal subtypes from images, relying on XGBoost for that step—a limitation that future multi-class image heads could address. The publicly sourced ultrasound images may also not represent the diversity of scanner manufacturers, imaging protocols, and ethnic populations encountered in global practice. Finally, the system currently lacks formal uncertainty quantification beyond the sigmoid probability output, which means it cannot reliably flag low-confidence predictions for human review.

XI. Future Work

The most pressing next step is prospective clinical validation in collaboration with obstetric departments at tertiary hospitals, benchmarking system outputs against specialist diagnoses on de-identified patient cohorts. Integrating Grad-CAM saliency maps would produce spatial attention overlays on ultrasound images, giving clinicians a visual explanation of what drives each prediction and building trust in the system's outputs. A model ensemble—stacking EfficientNetB0 with MobileNetV3 and DenseNet-121 predictions—could reduce variance for edge-case classifications.

The underlying framework architecture is also transferable beyond chromosomal aneuploidies to structural anomalies including neural tube defects, cardiac septal defects, and fetal growth restriction. Integration with hospital electronic health record systems via HL7 FHIR APIs would embed the tool seamlessly into existing clinical workflows. Longer-term, federated learning protocols could allow multi-institutional model training while keeping sensitive patient data local, directly addressing the privacy and regulatory barriers that currently limit cross-hospital collaboration.

XII. Conclusion

This paper has described a hybrid AI framework for prenatal fetal health screening that joins an EfficientNetB0 CNN with an XGBoost classifier in a sequential decision fusion pipeline. The system directly addresses a meaningful unmet need by offering automated, interpretable, and accessible screening for chromosomal aneuploidies from fetal ultrasound images and maternal clinical data. Evaluated on curated image and synthetic tabular datasets, the framework achieved 97.2% CNN binary classification accuracy and 94.6% XGBoost multi-class accuracy, outperforming ResNet-50 and VGG-16 baselines and comparing favorably with related systems in the recent literature. The Streamlit deployment delivers real-time predictions with sub-second latency on consumer-grade hardware, with no specialist infrastructure required.

We see this architecture as a scalable foundation for AI-assisted prenatal screening, with immediate value in resource-limited settings and a clear path toward progressive clinical validation and expansion. Once rigorously validated in prospective studies, tools of this kind have real potential to reduce the global burden of undetected congenital abnormalities.

References

[1] World Health Organization, “Congenital anomalies,” WHO Fact Sheet, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/congenital-anomalies

[2] L. S. Chitty, “Prenatal screening for chromosomal abnormality by ultrasonography,” Prenatal Diagnosis, vol. 15, no. 13, pp. 1241–1252, 1995.

[3] C. H. Gravholt, “Clinical practice in Turner syndrome,” Nature Clinical Practice Endocrinology & Metabolism, vol. 1, no. 1, pp. 41–52, 2005.

[4] A. Rajpurkar et al., “CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning,” arXiv:1711.05225, 2017.

[5] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2016, pp. 785–794.

[6] Y. Chen et al., “SKGC: A general semantic-level knowledge guided classification framework for fetal congenital heart disease,” IEEE Trans. Med. Imaging, 2023.

[7] H. Zhang et al., “A coarse-fine collaborative learning model for three-vessel segmentation in fetal cardiac ultrasound images,” Comput. Biol. Med., vol. 152, 2023.

[8] A. Alqahtani et al., “Artificial intelligence techniques for detection of congenital diseases: A systematic review,” IEEE Access, vol. 11, 2023.

[9] A. Rajendran and M. Jayabalan, “The smart earlier prediction of congenital heart disease in pregnancy using deep learning model,” Diagnostics, vol. 12, no. 8, p. 1875, 2022.

[10] J. Pinto et al., “Benchmarking deep learning architectures for fetal ultrasound plane classification,” Med. Image Anal., vol. 84, 2023.

[11] M. Burgos-Artizzu et al., “Fetal planes DB: A multi-centre volume dataset for ultrasound fetal brain segmentation,” Sci. Data, vol. 7, no. 1, pp. 1–11, 2020.

[12] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. ICML, 2019.

[13] J. Yosinski et al., “How transferable are features in deep neural networks?” Adv. Neural Inf. Process. Syst., vol. 27, 2014.