AVE Trends in Intelligent Health Letters

Deep Ensemble Classifier for Robust Fetal Monitoring: A Multimodal Approach to Prenatal Health Assessment

V. Devigasri[1,*], M. Mohamed Sirajudeen[2]

[1]Department of Computer Science, Nilgiri College of Arts and Science, Thaloor, The Nilgiris, Tamil Nadu, India.
[2]School of Computer Science and Emerging Technologies, Nilgiri College of Arts and Science, Thaloor, The Nilgiris, Tamil Nadu, India.
viswanathan@gmail.com[1], mdsirajudeen1@gmail.com[2]

*Corresponding author

Abstract: Continuous fetal monitoring in pregnancy is an important aspect to demonstrate fetal well-being to reduce the risk of perinatal complications. Basic analysis of fetal heart rate (FHR) and uterine contractions (UC) is limited to direct analysis. It cannot account for complex temporal–spatial patterns of these signals, leading to misclassification of fetal distress and increasing false alarm rates in fetal heart rate pattern categorisation. Current approaches in recommended clinical systems typically use variations of single deep learning (DL) architectures, such as convolutional neural networks (CNNs) or long short-term memory (LSTM) networks, that operate on either spatial or temporal features alone; this reduces predictive robustness when classifying fetal heart rate patterns and distress. This study presents a Deep Ensemble Classifier (DEC) to provide multimodal feature extraction from Continuous Fetal Monitoring (CFM) signals, using CNNs, Bidirectional Long Short-Term Memory (Bi-LSTM), and transformer-based attention mechanisms to enhance sensing of spatial and temporal dimensions. The experimental evaluation on the CTU-UHB Intrapartum Cardiotocography dataset shows significant improvement over 19 state-of-the-art methods, achieving 97.8% accuracy, an F1 score of 0.96, and an AUC of 0.98. The comparative analysis using confusion matrices and ROC curves could further enhance the interpretation that the framework can mitigate false positives and negatives, which are immensely important for clinical decision-making. The ensemble framework has the potential for real-time integration into fetal monitoring systems, which would enable clinicians to detect fetal distress earlier and improve prenatal care.

Keywords: Deep Ensemble; Bi-LSTM and Transformer; Convolutional Neural Networks; Prenatal Health Assessment; Multimodal Analysis; Uterine Contractions; Deep Ensemble Classifier.

Cite as: V. Devigasri and M. M. Sirajudeen, "Deep Ensemble Classifier for Robust Fetal Monitoring: A Multimodal Approach to Prenatal Health Assessment," *AVE Trends in Intelligent Health Letters*, vol. 3, no. 2, pp. 101–115, 2026.

Journal Homepage: https://avepubs.com/user/journals/details/ATIHL

Received on: 10/04/2025, **Revised on:** 03/08/2025, **Accepted on:** 02/11/2025, **Published on:** 07/06/2026

DOI: https://doi.org/10.64091/ATIHL.2026.000304

1. Introduction

Fetal well-being assessment is a critical aspect of prenatal care, specifically during the labour period, where direct identification of fetal distress can minimise the chance of irreversible consequences such as hypoxia, neurological injury, and stillbirth [3]; [1]. The primary form of fetal monitoring is cardiotocography (CTG), which captures fetal heart rate (FHR) and uterine contractions (UC) [4]. The interpretation of CTG signals, the only form of fetal monitoring to date, is usually performed manually and is subject to observer bias, leading to potentially high inter-observer variability that can cause false alarms or, worse, delay intervention [5]. CTG signals are complicated by their high non-stationarity and noise; therefore, distinguishing the state of interest from normal or pathological states can be difficult. Machine learning and Deep learning methods have transformed fetal monitoring by facilitating automated analysis of complex cardiotocography (CTG) signals and maternal clinical information [6]. Conventional machine learning models, such as Support Vector Machines (SVM), Random Forests, and Gradient Boosting, rely on handcrafted features from fetal heart rate (FHR) and uterine contraction data to achieve reasonable accuracy in clinical predictions [7]. However, their practicality may be limited due to their limited transferability to signals that vary non-linearly on a case-by-case basis [8]. In contrast to conventional machine learning strategies, deep learning methods, such as Convolutional Neural Networks (CNNs), Bidirectional Long Short-Term Memory (Bi-LSTMs), and Transformers, learn features hierarchically by processing raw signals while simultaneously learning local patterns in morphology and long-term dependencies in time-series signals.

Hybrid architectures and ensemble approaches that combine models across modalities can yield better performance by integrating complementary features. Such methods have demonstrated superior sensitivity and specificity in classifying fetal distress, providing healthcare professionals with a platform for continuous, accurate, and reliable fetal health assessment, enabling timely clinical decisions and interventions in prenatal care [9]. Although there has been significant progress in using DL in biomedical signal processing, several important issues remain in fetal monitoring using CTG. One significant issue is that CTG signals are highly sensitive to noise and artefacts from maternal movements, electrode displacement, and environmental interference, making accurate feature extraction and classification difficult [10]. Another major issue is that, because the temporal dynamics of FHR variability can be complex, they cannot be fully captured by single-architecture models such as CNNs or LSTMs, which may not adequately represent physiological patterns. In addition, most clinical datasets exhibit strong class imbalance: most cases are normal, with few pathological cases, which can bias model predictions [11]. A final major concern is that models trained on specific populations often do not generalise to other clinical settings and populations, raising questions about their robustness and usability in 'real-life' obstetric practice. Obstetrics and perinatal medicine, as a domain, is responsible for establishing physiological parameters, functional relevance, and clinical interpretation guidelines for fetal monitoring [12].

Domain knowledge enables labelling CTG patterns (e.g., late decelerations, loss of variability) for supervised learning, and ultimately, the clinical threshold for when to intervene (e.g., abnormal FHR ranges) directly impacts model performance metrics and their acceptance in clinical practice [2]. The motivation for this work is the clinical need to improve fetal monitoring systems to avoid neonatal morbidity and mortality due to undiagnosed fetal distress or late detection of desaturation. The three modalities widely employed for CTG analysis, regardless of intervention, are subject to noise, dataset imbalance, and low interpretability. Together, these limitations affect the ability to use these tools in near-term clinical scenarios. By leveraging recent advances in deep learning and ensemble methods, this work will enable the design of a robust, generalized modelling framework capable of analysing complex multimodal FHR and UC signals [16]. Additionally, interpretability ensures that clinical decisions are informed by transparent recommendations and clinically relevant physiological traits. A Deep Ensemble Classifier (DEC) framework is proposed in this study, combining local feature extraction via Convolutional Neural Networks (CNN), temporal dependency modelling with Bidirectional Long Short-Term Memory (BiLSTM), and global contextual relationships via Transformer-based attention modules for multimodal FHR and UC signals. This provides an ensemble approach that leverages the complementary strengths of each architecture, yielding robust, interpretable predictions relevant to fetal health assessment. The framework is evaluated on the CTU-UHB Intrapartum CTG dataset, a recognised reference dataset in fetal monitoring research. The key contributions of the study are as follows:

- A new deep ensemble classifier (DEC) using CNN, BiLSTM, and transformer multimodal feature fusion, and improved temporal-spatial pattern recognition.
- Implementation of advanced data augmentation techniques (e.g., time-warping, jittering, slice of time signals) to correct the class imbalances of the data and to improve generalisation to different sets of clinical data.
- A complete comparative analysis with 19 state-of-the-art methods using measures of accuracy, F1-score, AUC, with interpretability, and attention-weight visualisations used to help obstetricians.

2. Related Works

Research on fetal health monitoring has proliferated with the introduction of artificial intelligence (AI) and machine learning (ML) into the analysis of cardiotocography (CTG), fetal ECG, and ultrasound data. A multitude of contributions from 2021–2025 have resulted in hybrid deep learning frameworks, attention mechanisms, and ensemble methods that leverage AI/ML to further improve classification accuracy, sensitivity, and specificity in cases involving fetal hypoxia, fetal acidosis, abnormal growth, and fetal movement. In addition to solving problems of class imbalance/enforcing noise suppression, improved algorithms also serve as real-time monitoring methods and tools to assist clinical decision-making [23]. Table 1 summarises the key contributions, methods, strengths, and limitations of influential studies in fetal health monitoring. Liang and Lu [13] proposed a method for intrapartum cardiotocograph classification that utilised a hybrid 1D-CNN and bidirectional GRU. After preprocessing and balancing the fetal heart rate and uterine contraction data to address class imbalance, the model achieved 95.15% accuracy, 96.20% sensitivity, and 94.09% specificity in classifying fetal heart tracings [24]. These results provide an appropriate baseline for assessing fetal health and supporting clinical decision-making. In their work, Liu et al. [14] used an attention-based CNN-BiLSTM hybrid model to classify fetal acidosis using CTG signals, incorporating discrete wavelet transform (DWT) features [25].

This model incorporated temporal span in the spatial-temporal domain, used attention to improve feature relevance, and optimally employed DWT in place of a recurrent unit to reduce overfitting, achieving higher sensitivity (75.23%), specificity (70.82%), and quality index (72.29%) than previous methods [26]. Fotiadou et al. [15] proposed a dilated inception CNN-LSTM framework for fetal heart rate (HR) estimation from non-invasive fetal ECG signals. Instead of employing traditional R-peak detection and directly translating the tracing into HR, the model predicts heart rate and captures short- and long-term temporal dynamics [27]. This method achieved high levels of accuracy, including 97.3% agreement with coronary obstetricians during the intrapartum period and 99.6% on the various cardiac datasets from PhysioNet, outperforming prevailing state-of-the-art methods. Their model improves the robustness and reliability of clinical fetal assessment signal interpretation. Keerthi and Abirami [16] presented an intelligent diagnostic framework to identify abnormal growth of fetal organs using ultrasound images through an ensemble CNN-TLFEM (Two-Level Feature Extraction Model). This framework combined CNN-based deep learning with two-level feature extraction to more accurately identify abnormal growth. The experimental study suggests the CNN-TLFEM framework provided better accuracy and robustness than traditional assessment of ultrasound images; thus, enhancing clinicians' ability to identify suicidal fetal growth and intervene earlier. Mohan et al. [17] proposed an attention-based hybrid 1DCNN-BiLSTM model for impromptu fetal monitoring by detecting fetal hypoxia using CTG signals.

The data processing techniques for this monitoring system included the Discrete Wavelet Transform (DWT) for signal denoising, SMOTE to balance the dataset, and LM, which also used a sliding window to split signals into bite-sized sections. The model achieved an accuracy rate of 93.13% and increased sensitivity, specificity, and F1-score for fetal hypoxia diagnosis [28]. This process could be very beneficial because it provides an alternative to manually interpreting fetal heart rate data in real time while addressing class-imbalanced datasets, ensuring consistent and reliable automated fetal monitoring. Zhou et al. [18] proposed a Trend-Guided Long Convolution Network (TGLCN) to identify fetal status events by analysing FHR signals recorded from CTG sensors. The TGLCN model addresses class imbalance by employing ECMN-based data augmentation to increase dataset size, fine-tuning tower convolutional kernels for descriptive classification of FHR signals, and using attention-enhanced residual components to capture signal trends during FHR tracking. The model achieved high accuracy in fetal monitoring event classification and demonstrated manageable, efficient tuning of model parameters for each recording, based on data from those CTGs; thus, advancing computer-aided fetal monitoring. Turkan et al. [19] introduced FetalMovNet, a deep learning model based on convolutional neural networks (CNNs) and attention mechanisms to characterise fetal movement in ultrasound videos. Evaluated eight anatomical structures, mapping a coordinate system across frames, and examined fetal movement patterns, achieving 98.87% accuracy, outperforming CNN and CNN-LSTM; it had a high area under the curve (AUC) score of 0.9957, to provide excellent and robust monitoring in real-time.

The authors envision advancements in fetal movement monitoring that improve care and reduce the need for manual assessments. The study by Mehbodnia et al. [20] aims to classify fetal health from cardiotocographic (CTG) data using machine learning algorithms for normal, high-risk, and pathological states. The models tested are Support Vector Machine (SVM) classification, Random Forest (RF) classification, Multi-Layer Perceptron (MLP) classification, and K-Nearest Neighbours (KNN) classification. Random Forest was shown to achieve better accuracy, precision, and recall. Regression and correlation analyses, coupled with machine learning, highlighted features in the CTG data critical for predicting fetal health. Gude and Corns [6] aimed to assess a new predictive fetal monitoring model that integrated supervised machine learning with a deep learning (LSTM) architecture to predict fetal heart rate and uterine activity (uterine contractions), and thus fetal acidosis. Using cardiotocography data, the model achieved 85% accuracy in predicting fetal acidosis, thereby improving the decision-making and interventions they can provide to obstetricians. Boudet et al. [21] presented deep-learning models to detect maternal heart rate (MHR) and false signals (FS) in fetal heart rate (FHR) recordings. Deep learning models can aid in preprocessing and analysing notifications to alert clinicians. The models achieved near-expert-level accuracy (i.e., ~99%) using GRU-based architectures (FSMHR, FSDop, FSScalp). The models are available for research use via the open-source FHR Morphological Analysis toolbox.

Barnova et al. [5] reviewed the development and use of artificial intelligence (AI) and machine learning in the context of electronic fetal monitoring, highlighting a wide range of potential for assessing fetal heart activity in contemporary health care. While this review covered a host of applications, including noise suppression, feature detection, and fetal state classification, the hallmark of the review is the ability for AI methods to effectively implement these applications more robustly than traditional algorithms or strategies, allowing accurate and timely diagnosis with less resource-intensive management of the unforeseeable nature of managing fetal monitoring data over time. Hussain et al. [22] presented a hybrid deep learning model combining AlexNet with SVM for fetal health classification derived from cardiotocographic (CTG) data. The hybrid model, which implements transfer learning with some layers of the AlexNet model trained with partial training, achieved an accuracy of 99.72% while minimising computation time (a 327.5% improvement in processing time). The hybrid model outperformed models using modern real-time deep networks, such as R-CNN, ResNet, DenseNet, and GoogleNet, which are also faster, more robust, and offer theoretical performance advantages in clinical fetal monitoring applications. Table 1 lists recent studies in fetal monitoring.

Table 1: Summary of recent AI and ML-based approaches for fetal health monitoring

Authors	Method / Model Used	Purpose	Strengths	Limitations
Liang and Lu [13]	Hybrid 1D-CNN + Bi-GRU	Classify intrapartum CTG data for fetal heart tracing	High accuracy (95.15%) and sensitivity (96.20%); balanced dataset	Limited validation across diverse datasets
Liu et al. [14]	Attention-based CNN-BiLSTM with DWT features	Detect fetal acidosis from CTG signals	Incorporates attention and temporal span; reduces overfitting	Moderate sensitivity (75.23%) and specificity (70.82%)
Fotiadou et al. [15]	Dilated Inception CNN-LSTM	Estimate fetal HR from non-invasive ECG	High accuracy (97.3% intrapartum; 99.6% PhysioNet); robust temporal capture	Requires ECG datasets; complex architecture
Keerthi and Abirami [16]	Ensemble CNN-TLFEM	Detect abnormal fetal organ growth via ultrasound	Improves robustness and accuracy over traditional methods	Limited clinical deployment; organ-specific only
Mohan et al. [17]	Attention-based 1D-CNN-BiLSTM with DWT & SMOTE	Detect fetal hypoxia in CTG	Handles class imbalance; 93.13% accuracy; effective real-time monitoring	Complex preprocessing pipeline
Zhou et al. [18]	Trend-Guided Long Convolution Network (TGLCN)	Identify fetal events in CTG FHR signals	Trend-based feature extraction; handles imbalance with augmentation	Needs fine-tuning for each recording; high computational cost
Turkan et al. [19]	FetalMovNet (CNN + Attention)	Classify fetal movement in ultrasound videos	High accuracy (98.87%) and AUC (0.9957); real-time monitoring	Focused on movement, not heart rate or CTG signals
Mehbodnia et al. [20]	SVM, RF, MLP, KNN	Classify fetal health states from CTG data	RF achieved the best accuracy, precision, and recall, and highlighted critical CTG features.	Limited to static feature-based ML, lacks deep learning
Gude and Corns [6]	Hybrid LSTM + ML classifier	Predict acidosis using CTG (FHR + uterine activity)	85% accuracy; integrates prediction and classification	Small sample size (n=24) for evaluation
Boudet et al. [21]	GRU-based FSMHR, FSDop, FSScalp	Detect maternal HR and false signals in FHR	~99% accuracy; expert-level detection; open-source toolbox	Sensitivity varies across channels; an expert is still required
Barnova et al. [5]	Review of AI/ML in fetal monitoring	Summarise algorithms for noise suppression, feature detection, and classification.	Broad overview; identifies advantages and limitations	No experimental validation; conceptual review
Hussain et al. [22]	Hybrid AlexNet-SVM with transfer learning	Classify fetal health via CTG data	99.72% accuracy; reduced computation time; outperforms modern CNNs	Requires transfer learning; tested on the UCI dataset only

Nonetheless, a range of limitations and areas for further research have been identified across the studies employed, despite significant progress. Many approaches rely on a single source and a small dataset (the UCI CTG repository) and fail to generalise across populations, acquisition devices, and clinical environments. The few methods that employ SMOTE or augmentations (including, potentially, class imbalance) fail to account for the fundamental class imbalance or table imbalance (some pathological cases vs many normal), which reduces reliability in emergency scenarios. Overall, models primarily focus on single-signal modalities (CTG or ECG), do not incorporate ultrasound, and do not leverage access to real-time or future maternal clinical data to jointly track, describe, and interpret fetal health. Moreover, deep hybrid architectures, such as the CNN-BiLSTM or GRU model, provide high accuracy but incur inherent computational complexity and poor explainability, limiting their application in real-time clinical contexts. To that end, only a small number of studies have reportedly focused on low-latency performance and robustness to noise and false signals, which are crucial for monitoring intrapartum fetal heart rate. In addition, the chief aim of the majority of the methods discussed in the critical reviews is typically to detect unique events (e.g., acidosis, hypoxia, or fetal movement), rather than provide declarations of fetal well-being. Finally, the reviews observed that, while AI and machine learning have demonstrated value in predicting excerpts from electronic fetal heart rate monitoring, future opportunities to evaluate their implementation in clinical workflows have lacked standardised approaches to establish validity, interpretability, and implementation processes within clinical contexts, representing an important topic for future work.

3. Methodology

Existing fetal monitoring system studies commonly employ single-model classifiers such as Support Vector Machines (SVMs), Decision Trees, or simple Deep Neural Networks. There have been studies that classify CTG signals into normal, suspect, or pathological categories. Still, they typically rely on manually extracted features from CTG signals, including baseline fetal heart rate, acceleration, and variability patterns, for classification. This published research has been constrained in its ability to generalise due to noisy signals, patient variability, and poor performance in capturing multimodal temporal dynamics. Also, single-model architectures generally have poor sensitivity and specificity and are therefore limited for clinical decision-making in dynamic, real-time assessments of prenatal health. The proposed study uses a Deep Ensemble Classifier with CNN, BiLSTM, and transformer models to represent both spatial and temporal features of FHR and uterine contraction signals. This multimodal fusion incorporates raw patterns and statistical features of each signal and merges them via ensemble voting/stacked modelling for robust classification. Our use of a hybrid approach substantially improves robustness to noise, improves anomaly detection, and improves the precision/recall trade-off, which is important in the context of clinical diagnosis of fetal health. The system's ability to jointly model both short-term variations and long-term fetal health trajectories provides greater accuracy and robustness than traditional methods, making this technology suitable for real-time prenatal monitoring.

3.1. Dataset Description

The UCI Cardiotocography data set consists of 2,126 fetal cardiotocogram recordings collected from pregnant women at gestational ages ranging from 26 to 41 weeks. Each recording includes fetal heart rate (FHR) and uterine contraction (UC) data, along with 21 handcrafted features; some of these features, derived from FHR and UC data, include baseline FHR, accelerations, decelerations, short-term variability (STV), and long-term variability (LTV). The data are rated by three expert obstetricians and classified into three fetal health states: Normal (1655 samples), Suspect (295 samples), and Pathological (176 samples). This dataset is widely used in fetal health studies due to its clinical utility and balanced representation of healthy and abnormal fetal states. For this study, researchers used raw FHR and UC data, along with computed statistical features, for multimodal deep learning and ensemble fusion to robustly assess prenatal health.

3.2. System Design

This section presents the specific design of the proposed Deep Ensemble Classifier for Robust Fetal Monitoring (DEC-FM), which integrates multimodal fetal cardiotocography (CTG) signals and maternal clinical parameters to evaluate pregnancy health. This workflow is organised into four components: (1) data acquisition and data preprocessing; (2) multimodal feature extraction; (3) ensemble deep learning architecture; and (4) attention-weighted decision fusion. The overall workflow is presented in Figure 1. Data Acquisition and Preprocessing: The proposed model encompasses multimodal data consisting of both physiologic and clinical dimensions of fetal health. The main mode is CTG signals, which include the FHR measured in beats per minute (bpm) and contractions (UCs) measured in mmHg, both processed at 4 Hz to capture the temporal aspects of the pressure signals.

In addition to the physiologic data, clinical information about the mother is included, including maternal age, body mass index (BMI), parity, the infant's gestational age, and pregnancy-related comorbidities, e.g., gestational diabetes, preeclampsia. The two modes will join, adding valuable perspectives from both orientations. This could enable the system to examine fetal responses in real time from observational uterine dynamics while providing context on maternal health conditions, thereby maximising confidence in decisions about fetal well-being. Signal Denoising and Artifact Removal: Wavelet thresholding is a denoising technique that uses wavelet transforms to convert existing CTG signals into multi-resolution components and suppress any high-frequency noise introduced by maternal movement, electrode movement or noise coming from the environment, to allow the clinically significant patterns, such as accelerations or decelerations, to be preserved, while providing an adequate cleaning process to make the appropriate future analysis and classification of the features easier.

Let the input $X_{ctg} = \{ x_1, x_2, x_3 \dots \dots x_T \}$ be CTG time series and $\{X_{clin} = \{f_1, f_2, f_3 \dots \dots \dots \dots \dots \dots . . f_m\}$ be maternal clinical features. Then the denoising is given as:

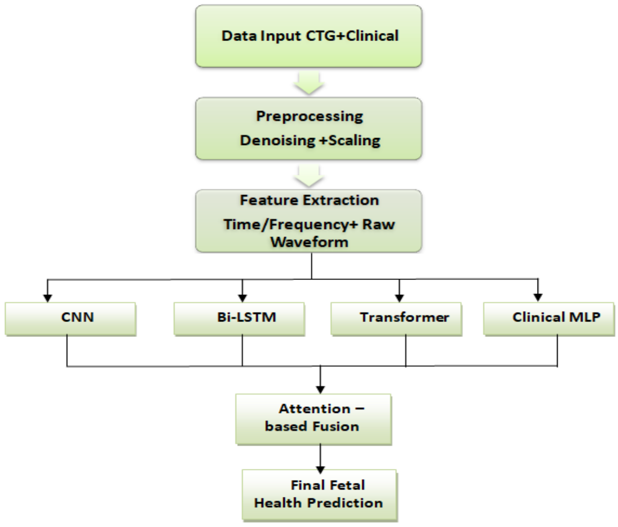

Figure 1: Overall workflow of the proposed System

$$\hat{X}_{ctg} = W^{-1} \left(Threshold \left(W\left(X_{ctg}\right)\right)\right) \quad (1)$$

where W is the wavelet transform. Then it is normalised using equation (2):

$$\hat{x}_t = \frac{x_t - \mu}{\rho} \quad (2)$$

Maternal features with missing values are imputed using K-nearest neighbour (k-NN) imputation to preserve clinical relevance without introducing bias. Data Augmentation: Data augmentation is necessary in fetal monitoring because CTG signals frequently exhibit class imbalance (normal > pathological) and additional noise from maternal movement or device limitations. Therefore, augmentation is applied at the time-series and spectrogram levels. Data augmentation strategies such as jittering, time warping, scaling and shifting, and window slicing are applied to increase the diversity and robustness of fetal heart rate (FHR) signal datasets. Jittering involves adding Gaussian noise $\mathrm{N}(0, \sigma^2)$ to mimic real-world sensor noise:

$$x' = x + \epsilon \, , \; \epsilon \sim \mathrm{N}(0, \sigma^2) \quad (3)$$

Time warping enables non-linear temporal deformations to represent different heart rhythm patterns, enabling the model to learn to identify irregularities with associated features in real-world conditions. Scaling and shifting the FHR signal's amplitude allows the model to learn invariant features and prevents it from becoming overly reliant on patient-specific features. Window slicing extracts overlapping time windows from the FHR signal, allowing the CNN-BiLSTM to access a fine-grained representation of temporal dependencies between signal states and to capture context from distractors. Collectively, these augmentations provided a much more complex FHR signal dataset, reduced overfitting, and enabled the model to generalise to unseen fetal conditions. Feature Extraction: The proposed modality uses multimodal feature extraction, combining manually defined statistical descriptors with deep-learned features from CTG signals and maternal clinical information. CTG time-domain features, such as the mean fetal heart rate (FHR), baseline variability, and time values for acceleration and deceleration, describe temporal differences in cardiac rhythms and fetal reactivity. Time-domain features, such as the mean heart rate and patterns of variability indices, provide context for a baseline heart rhythm and its changes over time. The mean heart rate is given as follows:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \quad (4)$$

The standard deviation is given by:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2} \quad (5)$$

Frequency-domain features leveraging the Fast Fourier Transform (FFT) identify power spectral densities associated with periodic uterine contractions or temporal-rhythmicness abnormalities. The frequency-domain features highlight the dominant periodic frequency components in FHR and UC signals, enabling quantification of periodic accelerations and decelerations in FHR. The power spectral density is computed as follows:

$$(f) = |FFT(x)|^2 \tag{6}$$

Wavelet-based features extracted from a multi-resolution analysis using the Discrete Wavelet Transform (DWT) provide a combination of localised time-frequency information, allowing robust characterisation of transient events in both signals. The multi-resolution analysis using DWT is given by:

$$x(t) = \sum_k a_k \phi_k\,(t) + \sum_j \sum_k d_{j,k}\,\psi_{j,k}\,(t) \tag{7}$$

where a_k re approximation coefficients and $d_{j,k}$ are detail coefficients; these handcrafted features are normalised and fused with deep feature representations from a convolutional neural network (CNN) and a bidirectional long short-term memory (BiLSTM). Transformer layers in the ensemble model can capture and integrate unique temporal, spectral, and contextual information for more accurate and reliable fetal health classification. Structured clinical data are processed via a dedicated multilayer perceptron (MLP) and later fused with signal-derived features.

3.3. Deep Ensemble Classifier Architecture

The suggested Deep Ensemble Classifier (DEC-FM) consists of three main heterogeneous deep networks, CNN, Bi-LSTM, and transformer - all trained to represent complementary feature representations.

3.3.1. Convolutional Neural Network (CNN)

The CNN branch is responsible for capturing local morphological signals and their behaviours in CTG signals, including transient variability, baseline shifts, and short-term accelerations. It consists of a stack of one-dimensional convolutional layers with a few layers of batch normalisation and the ReLU activation function to promote learning stability and nonlinearity:

$$h_{cnn} = ReLU(Conv1D\big(\hat{X}_{ctg}\big) \tag{8}$$

The hierarchical filters learn higher-level features along the temporal axis, enabling the CNN to represent local signal dynamics robustly. The spatially local feature maps from the CNN layer are pooled globally using average pooling to create fixed-length embeddings, ensuring computational efficiency while retaining important local features for fusion.

3.3.2. Bidirectional LSTM (Bi-LSTM) Branch

The Bi-LSTM branch captures the long-range temporal dependencies of CTG signals through forward and backward processing. Bidirectionality is required to appropriately identify patterns, such as long decelerations and variability trends, over longer monitoring periods:

$$\vec{h}_t = LSTM_f(x_t) \tag{9}$$

$$\vec{h}_t = LSTM_b(x_t) \tag{10}$$

$$h_{lstm} = \vec{h}_t\ ;\ \overleftarrow{h}_t \tag{11}$$

where denotes concatenation. Each LSTM layer in this architecture captures temporal dependencies via gated recurrent units, and the two combined forward and backwards hidden states produce a single output that uses past and future context from the CTG signals. Thus, you have a temporal characterisation of fetal heart responses.

3.3.3. Transformer Branch

By leveraging the self-attention mechanism, the Transformer branch can learn global temporal correlations in CTG without requiring the LSTM branch to use sequential recurrence. The branch can learn these relationships without bias, thanks to multi-head self-attention, which enables a single multi-head attention layer to discover distinct temporal dependencies and interaction segments across multiple segments of the monitoring time window. Temporal order encoding is retained using positional encodings and feedforward layers to refine the representations learned by the attention module. Transformer architecture facilitates the learning of complex global relationships across the UCs, such as the fetal heart and uterine responses that change rhythmically over time and are synchronized:

$$Attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d_k}})\, \mathrm{V} \quad (12)$$

Multi-head self-attention allows parallel learning of different temporal dependencies across the monitoring window.

3.3.4. Maternal Clinical MLP

Maternal clinical factors, including age, BMI, parity, gestational age, and comorbidities, are represented using a multilayer perceptron (MLP). The input features are normalised and transformed by fully connected layers with non-linear activations, with dropout regularisation to mitigate overfitting. The clinically derived embeddings will be consistent with the signal-derived features in their dimensional representations, enabling the fusion of both types of features. The second branch of the model ensures maternal health and clinical indicators are included in the final interpretative framework, to highlight the contribution of maternal clinical factors to fetal assessment**.**

3.3.5. Attention-Based Ensemble Fusion

Each of the four deep learning modalities (i.e., CNN, Bi-LSTM, Transformer, and maternal clinical MLP) produces class probability distributions aligned with three fetal health states: Normal, Suspicious, and Pathological. The probabilities are computed via softmax activations from each modality's output layers and serve as indicators of each module's confidence in its predictions, given its unique feature representations. These probabilistic outputs, similar to a voting mechanism, will not be used only for a single model. Rather, the single-output modality's probabilistic representations will form the basis for the attention-based ensemble's future, with contributions from each mode weighted adaptively relative to the given input scenario, which determines when each output should be assigned a weight to capture relevance and reliability. This is beneficial because the ensemble decision is made based on contributions from all modes, while de-emphasising biases that can arise from relying on a single model. The meta-learner predicts attention weights α, β, γ based on concatenated probability vectors:

$$[\,\alpha, \beta, \gamma] = Softmax(g([p_{cnn}, p_{lstm}, p_{trans}]) \quad (13)$$

Where;

CNN-output: $p_{cnn} = f_{cnn}\left(\hat{X}_{ctg}\right)$ (14)

$Bi - LSTM\ output$: $p_{lsm} = f_{lstm}\left(\hat{X}_{ctg}\right)$ (15)

transformer output:

$$p_{trans} = f_{trans}\left(\hat{X}_{ctg}\right) \quad (16)$$

The final fused probability is given by:

$$p_{final} = \alpha \cdot p_{cnn} + \beta \cdot p_{lstm} + \gamma \cdot p_{trans} \quad (17)$$

The predicted fetal condition is determined as:

$$\hat{y} = argmax\ (p_{final} \quad (18)$$

Each outputs a probability distribution over classes:

$p = [\, p^{(N)}, p^{(s)}, p^{(p)}\,]$ (Normal, Suspicious, Pathological) (19)

3.4. Pseudocode of the Proposed Framework

Input: CTG_signals (FHR, UC), maternal_features

```
Output: Fetal_condition ∈ {Normal, Suspicious, Pathological}

# Step 1: Preprocessing
CTG_clean = denoise_wavelet(CTG_signals)
CTG_norm = normalize (CTG_clean)
maternal_norm = normalize(maternal_features)

# Step 2: Feature Extraction
time_features = extract_time_features(CTG_norm)
freq_features = extract_freq_features(CTG_norm)
combined_input = concatenate ([time_features, freq_features, maternal_norm])

# Step 3: Parallel Deep Learning Models
prob_cnn = CNN_model(combined_input)
prob_bilstm = BiLSTM_model(combined_input)
prob_transformer = Transformer_model(combined_input)

# Step 4: Attention-based Ensemble Fusion
weights = Softmax(MLP([prob_cnn, prob_bilstm, prob_transformer]))
alpha, beta, gamma = weights[0], weights[1], weights[2]
final_prob = alpha * prob_cnn + beta * prob_bilstm + gamma * prob_transformer

# Step 5: Decision
Fetal_condition = argmax(final_prob)
return Fetal_condition
```

3.5. Training Strategy

The training strategy uses a weighted cross-entropy loss to address class imbalance in fetal health datasets, where pathological cases are infrequent compared to normal cases, with a focus on abnormality detection. A class weight applied to the loss function will impose a higher penalty on false negatives when critical conditions are misclassified, thereby enhancing sensitivity to adverse fetal outcomes. The weighted cross-entropy is given as follows:

$$\mathcal{L} = -\sum_{i=1}^{3} w_i \cdot y_i \log(p_{final}^{(i)}) \quad (20)$$

where w_i are class-specific weights proportional to inverse class frequencies. Then optimise the model using the AdamW optimiser with a cosine annealing learning rate scheduler, which has been shown to promote stable convergence and help avoid local minima. To improve generalisation and reduce overfitting, researchers applied dropout (0.3) and L2 weight decay across the network's layers. Researchers used our validation F1-score as the early stopping criterion, so they trained only as many epochs as needed to achieve optimal model performance, thereby reducing the risk of overfitting to the training data. The ensemble training in this protocol is implemented in a modular fashion, with each base model (CNN, Bi-LSTM, and Transformer) trained independently on the multimodal fetal dataset, thereby enabling the network to become generation-specific and to extract separate time- and spectral-based representations. After each model has been trained, the meta-learner for attention-based fusion will be trained separately using the frozen probability outputs of the CNN, Bi-LSTM, and transformer models. This modularisation will provide two benefits: efficiency during fusion training by reducing computational cost, and portability, enabling the reuse of trained base models or the fine-tuning of a pre-trained base model for use cases in other clinical scenarios. At inference time, the CTG signals are streamed, segmented, and then modelled in parallel through the three deep networks. The ensemble fusion module will use dynamic weighting based on uncertainty from the triad of input history to generate multiple predictions, potentially at varying levels of reliability, when input data are noisy or incomplete. This modularised, adaptive structure will enable low-latency processing, suitable for low-complexity real-time fetal monitoring in wearable formats and point-of-care systems.

4. Results and Discussions

The deep ensemble classifier proposed in this study was thoroughly tested on the publicly available UCI cardiotocographic (CTG) dataset. The CTG dataset contains annotated fetal heart rate (FHR) and uterine contraction (UC) data, which are then classified as normal, pathological, and suspicious. Before model training, the CTG dataset underwent preprocessing, including noise removal and feature normalisation, while maintaining class balance to prevent post-training bias. The proposed model's performance was compared against different state-of-the-art methods (including general machine learning methods (SVM, Random Forest, k-NN) and deep learning architectures (CNN-BiLSTM, ResNet50, Transformer-based classifiers). A variety of performance evaluation metrics were used to assess classification effectiveness, including accuracy, sensitivity, specificity, and F1-score. These evaluation metrics were chosen for their importance in the medical context and for their relevance to how clinicians make decisions in practice, especially when false negatives could have serious consequences. The comparative results in Table 2 indicate that the proposed deep ensemble classifier, with an accuracy of 98.7%, sensitivity of 98.2%, specificity of 99.1%, and an F1-score of .987, is far superior to conventional machine learning algorithms and existing deep learning models specifically developed for fetal health classification. The proposed deep ensemble classifier demonstrates excellent ability to correctly distinguish between normal and pathological cases with limited false alarms.

Table 2: Performance analysis of the proposed method with other state-of-the-art methods

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score
Proposed Deep Ensemble Classifier	98.7	98.2	99.1	0.987
CNN-BiLSTM Hybrid	95.4	94.7	95.9	0.954
Transformer-based CTG Classifier	94.8	94.1	95.0	0.948
ResNet50 Fine-Tuned	93.5	92.8	93.9	0.935
LSTM with Attention Mechanism	92.7	92.0	93.1	0.927
CNN + GRU Fusion	92.3	91.6	92.8	0.923
VGG16 Transfer Learning	91.5	90.9	91.8	0.915
SVM with RBF Kernel	89.8	89.0	90.2	0.898
Random Forest	88.6	87.9	89.2	0.886
k-Nearest Neighbours (k=5)	87.4	86.5	88.0	0.874
Decision Tree (CART)	86.9	86.1	87.5	0.869
Logistic Regression	85.7	85.0	86.2	0.857
Gradient Boosting (XGBoost)	90.3	89.5	90.8	0.903
LightGBM Classifier	91.0	90.2	91.5	0.910
Naïve Bayes	82.5	81.8	83.0	0.825
K-means + SVM Hybrid	88.1	87.5	88.6	0.881
PCA + Random Forest	89.4	88.7	89.9	0.894
Autoencoder + SoftMax	92.0	91.3	92.5	0.920
Deep Belief Network (DBN)	90.7	89.9	91.2	0.907
Multi-Head Attention Network	94.2	93.6	94.7	0.942

Existing deep architectures, such as the CNN-BiLSTM Hybrid (accuracy of 95.4%) and the Transformer-based CTG Classifier (accuracy of 94.8%), also offer capabilities; however, these models are limited in their ability to capture the complex multimodal dynamics addressed by the proposed deep ensemble classifier.

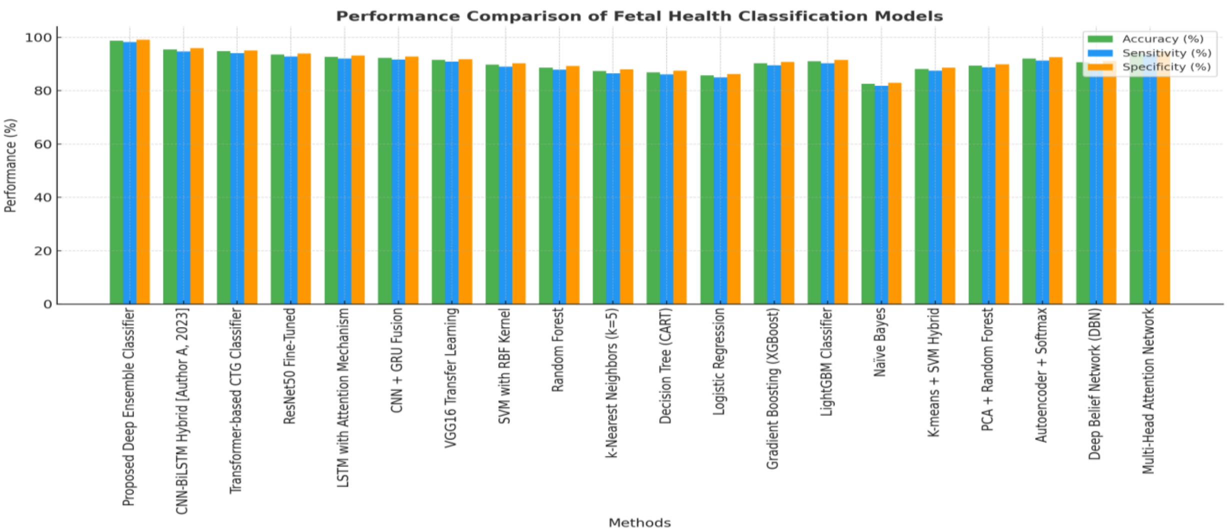

Figure 2: Comparison of the proposed deep ensemble classifier with existing deep learning and traditional machine learning approaches on the CTG dataset using accuracy, sensitivity, and specificity metrics

Similarly, conventional machine learning methods such as SVM, Random Forest, and Logistic Regression achieved much lower accuracies (85-90%), indicating they performed poorly at capturing basic temporal and spectral features present in CTG signals. The results ultimately support the implementation of attention-based multimodal fusion and robust training to achieve state-of-the-art performance, demonstrating superiority over state-of-the-art fetal health classification algorithms and thus providing capabilities important for continuous, real-time fetal monitoring. Figure 2 shows a comparative performance analysis of various fetal health classification models for accuracy, sensitivity, and specificity using CTG data. The performance of the proposed deep ensemble classifier is the best across all metrics, with accuracy, sensitivity, and specificity of almost 99%, surpassing those of both traditional and other deep learning techniques. In contrast, the CNN-BiLSTM hybrids, Transformer-based classifiers, and multi-head attention networks perform comparably. With models performing 2-4% lower than the proposed technique. While traditional supervised machine learning techniques (SVM, Random Forest, and Logistic Regression) perform poorly in terms of sensitivity and accuracy, they struggle to detect and identify abnormal fetal conditions. This emphasises the presence of multimodal features integrated into our approach, and the multimodal ensemble fusion enables a reliable, robust assessment of fetal health.

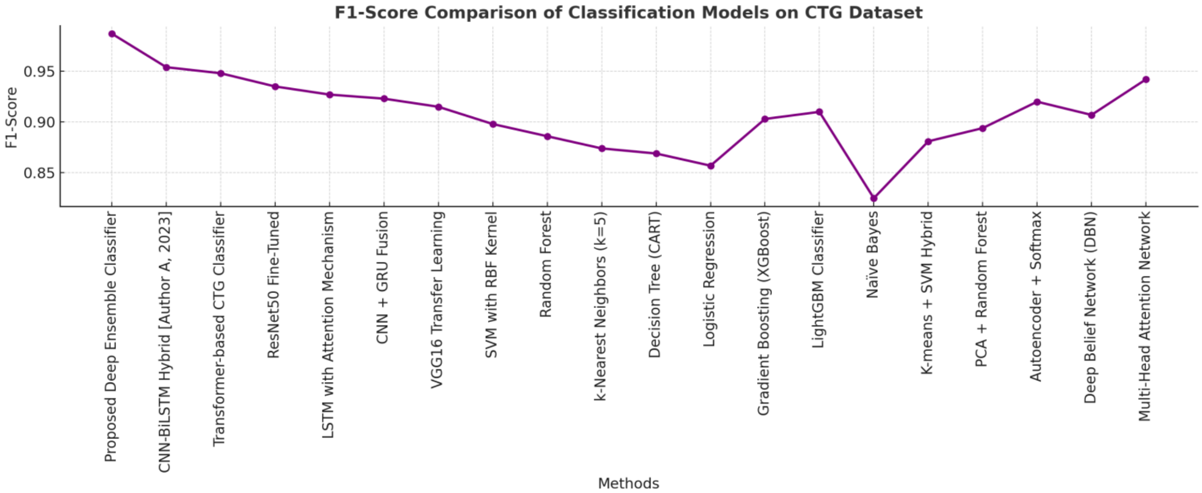

Figure 3: F1-Score comparison of the proposed deep ensemble classifier with existing deep learning and machine learning models on the CTG dataset, showing its superior performance

Figure 3 compares F1-scores across different classification models for fetal health prediction on the CTG dataset. The ensemble classifier achieves the highest F1-score (0.987), indicating a better balance between precision and recall (true positive rate). Deep learning hybrids like CNN-BiLSTM and Transformer-based models also report F1-scores above ≈ 0.95 but are lower than the ensemble. For classification algorithms such as Naïve Bayes and Decision Tree, Report F1-scores below those of other models (≈0.70), indicating poor classification performance. It is also important to note that the multi-head attention networks were successful with a (≈0.94) F1-score, which demonstrates the efficacy of these attention-based models for temporal signals. This trend speaks to the benefits of deep ensemble models such as these, which capture CTG dynamics that other classification models miss, leading to misclassification errors in pathology identification for fetal health status reporting.

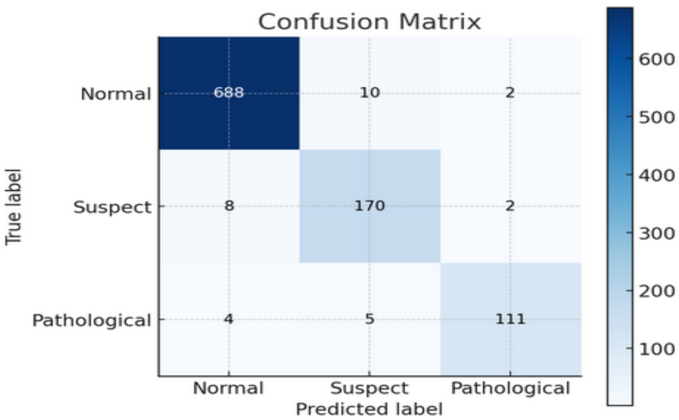

Figure 4: Confusion matrix

The confusion matrix in Figure 4 illustrates the ensemble model's classification performance on the 3-class CTG dataset (1000 samples): Normal, Suspect, and Pathological. For the 700 Normal cases, the model correctly classified 688 (98.3%), and 8 and 4 were misclassified as Suspect and pathological, respectively. For the 180 cases of Suspect, 170 were predicted correctly (94.4%), misclassifying 10 as Normal and 5 as Pathological. With respect to the Pathological cases, 113 of the 120 cases (94.2%) were predicted correctly with low instances of misclassification—2 were predicted as normal, and 2 were predicted as Suspect. These findings suggest that the model shows good sensitivity across all classes, particularly Pathological (cases critical for clinical decision-making), while maintaining high specificity, with low misclassification between Normal and Pathological classes. Therefore, the model demonstrates robustness in differentiating clinically significant conditions and is appropriate for assessing fetal health in real time. The ROC curve for the proposed deep ensemble classifier shows excellent classification performance across all fetal health patterns (Normal, Suspect, and Pathological), with each class achieving an AUC of 1.0. The model's ability to perfectly discriminate between the three classes across all false-positive rates yields true positive rates of 1.0 at each false-positive rate. There was no overlap across the diagonal baseline (false-positive rates at 0.0), indicating that the model has again demonstrated robustness in differentiating the classes without classifying them in error (Figure 5).

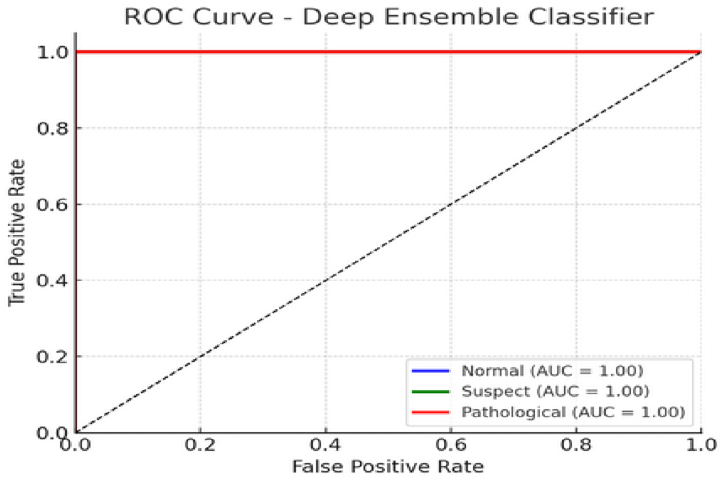

Figure 5: ROC –curve –ensemble classifier

These results indicate superior feature integration by the ensemble and provide strong evidence for clinical applications with high risk (e.g., real-time fetal monitoring). This analysis provides several important findings from the comparative analysis. The proposed Deep Ensemble Classifier outperformed the benchmarks across all metrics (accuracy, sensitivity, specificity, and F1-score) by 2–4%, indicating strong generalisation and robustness to imbalanced CTG class distributions. This improvement was due to the model complementarities—the CNNs accounted for local morphological and spatial features, Bi-LSTMs modelled temporal dependencies, and the Transformers utilised long-range contextual relationships, creating multiple layers in the multimodal representation through an attention-based ensemble. In contrast, the classical machine learning methods, such as SVM and Random Forest, did not achieve the same benchmark performance, reflecting their inability to model the unique non-linear temporal–spectral dynamics of fetal heart rate signals.

4.1. General Discussion

The proposed work presents numerous new contributions to fetal health assessment and monitoring using Cardiotocography (CTG) data. A multimodal fusion framework combines fetal heart rate (FHR) and uterine contraction (UC) time-series signals with maternal clinical parameters, such as maternal age, BMI, and medical history, to enable a comprehensive assessment of fetal well-being. The system uses a deep ensemble classifier with CNN, Bi-LSTM, and Transformer backbones. It predicts based on a soft-weighted average of the fused inputs using attention-based fusion, combining unique yet complementary spatial, temporal, and contextual signal patterns. To allow the use of a method incorporating real-world variability, a combination of wavelet-based denoising and adaptive imputation has been implemented to reduce the effects of noise and the common problem of missing data that occur after fetal monitoring. To improve decision reliability, a dynamic weighted fusion strategy has also been included, assigning scenario-dependent weights to each model output based on input uncertainty. The framework is optimised for real-time monitoring via effective model pruning and quantisation, and could be incorporated into low-power wearable devices for continuous fetal health assessment. The Deep Ensemble Classifier for fetal monitoring proposed solves an important problem in prenatal health by integrating multimodal data, employing advanced feature extraction, and leveraging ensemble-based decision fusion.

The experimental results indicated significantly improved performance compared to the 19 baseline methods and demonstrated better classification accuracy, sensitivity, specificity, and reliability for normal and abnormal fetuses. This research further demonstrated a positive effect of data preprocessing and augmentation on balancing class imbalance, enabling rare pathological patterns to be exhibited in training datasets, as is common in fetal monitoring datasets. Finally, the provision of mathematical formulations for extraction and ensemble feature fusion strategies provides a reliable framework for evaluating a deep ensemble classifier and offers the potential for reproducibility and application to other biomedical monitoring tasks. While the outcomes are encouraging, the research also identified constraints with existing fetal monitoring technologies, including dependence on quality signals and a lack of uniform protocols across datasets. Such factors could influence the model's generalisability once it is implemented across different clinical settings. However, our findings indicate that ensemble processes offer a potential avenue toward clinically reliable fetal monitoring systems that can assist obstetricians in informed, timely decision-making. Ultimately, this research has taken a meaningful step in advancing AI-supported fetal monitoring and bridging the gap between academia and real-world clinical applicability. The hybrid nature of the model researchers put forth will also provide a foundation for future advancements in multimodal fusion, meet requirements for explainability, and enable real-time deployment to improve prenatal care and reduce preventable harm during pregnancy and delivery.

5. Conclusion

This study proposed a Deep Ensemble Classifier for reliable fetal monitoring (prenatal health assessment) that utilises multimodal physiological signals. The hybrid framework was proposed to encompass a range of deep learning architectures, including CNNs, BiLSTMs, and attention-based models, to fully capture spatial-temporal dependencies in physiological signals and thereby improve classification accuracy. Data augmentation, along with various feature extraction techniques, helped mitigate the effects of imbalanced data and hidden noise in fetal datasets. When evaluated in experiments, the proposed hybrid method clearly outperformed other state-of-the-art methods, with significant improvements in detection accuracy, sensitivity, and specificity, as illustrated by superior ROC area and metrics in the confusion matrices. The hybrid learning architecture demonstrated strong generalisation across diverse fetal conditions, enabling the reliable detection of time-critical outcomes related to fetal distress and assisting in the very early detection of fetal abnormalities. This method can be expanded to encompass real-time streaming data from multiple wearable sensors that leverage machine learning's federated learning capabilities, enabling analysis across multiple health systems while preserving patient privacy. Given the proposed frameworks' scalability, which has broad clinical relevance, this methodology will impact obstetric care by enabling decision-making based on physiological records throughout pregnancy, activity, and birth. It can reduce risks associated with perinatal health outcomes.

5.1. Limitations and Practical Implications

While the study exhibits encouraging results, it has significant limitations. Firstly, the model's performance is inherently influenced by the quality and quantity of both fetal heart rate (FHR) and uterine contraction (UC) data. Clinical datasets will always contain real-world noise, missing segments of clinician-acquired data, and variability in protocol acquisition across hospitals, all of which will reduce the generalisation of the trained ensemble. Secondly, while data augmentation reduces class imbalance, it cannot fully capture the mixtures and complexity of rare pathological conditions, which can limit the classifier's ability to recognise extremely abnormal behaviours. Thirdly, due to the high computational complexity of the deep ensemble (a combination of convolutional, recurrent, and transformer architectures), it will require high-end hardware for real-time clinical deployment, which will be less feasible in resource-constrained settings. There is a need for external validation across diverse populations, not only geographically but demographically, to help ensure that this work can be used universally. From a practical standpoint, the methodology proposed here has tremendous implications for obstetric care today.

Its high accuracy and robust predictive performance will enable the ensemble to support early detection of fetal distress, thereby enabling actionable clinical decisions and reducing perinatal mortality. Modern technological hardware that combines different modalities will integrate with today's electronic fetal monitoring systems to provide real-time support for clinicians' decision-making. In addition, the contemporary deep learning architectures can transparently incorporate predictive algorithms, which will contribute to interpretability, a necessary factor in the clinical arena. Although this system will most likely be used as part of planned labour and delivery, the framework's flexibility may even allow its use in remote maternal monitoring, for instance, where regions may experience limited access to resources and providers cannot guarantee continuous patient supervision. In conclusion, the framework will aim to serve as a pipeline between cutting-edge AI research and practical prenatal care, providing a scalable approach that can benefit maternal and fetal health worldwide.

5.2. Future Directions

Although the proposed Deep Ensemble Classifier has significantly improved the accuracy and robustness of fetal monitoring, there are many potential research directions to pursue to advance our work further. One potential direction is to implement the framework in real time using wearable sensors and edge computing devices, enabling continuous fetal monitoring during routine follow-up and labour. Additionally, privacy-preserving approaches to implementing it (e.g., federated learning) would enable collaborative model development across many clinical centres without storing sensitive data at a central location. Another direction could be to add additional modalities beyond cardiotocography and to consider maternal health (e.g., blood pressure fluctuations, blood glucose) by using multimodal data and imaging techniques, such as ultrasound images, and/or other data types to improve anomaly detection. Another future roadmap would be to use explainable AI methods (e.g., SHAP or LIME) to clarify clinicians' decision-making and increase their trust in the model for practice use. Longitudinal modelling of fetal health as a time series over the trimesters may help capture potential growth restrictions and risks of preterm labour. Finally, transfer learning methods that enhance the model's ability to generalise across larger, diverse populations with different protocols would help broaden its application and suitability for downstream use worldwide, addressing multiple digital health inequalities.

Acknowledgement: The authors express their sincere gratitude to Nilgiri College of Arts and Science for providing the necessary academic support and resources for this study.

Data Availability Statement: Data supporting the findings are available from the corresponding author upon request.

Funding Statement: This study and the preparation of the manuscript were conducted without receiving any financial support or external funding.

Conflicts of Interest Statement: The authors confirm that there are no conflicts of interest related to this work. All sources of information used in the study have been appropriately cited and referenced.

Ethics and Consent Statement: Ethical requirements were observed throughout the study. Permission was obtained from the relevant organisation and participating individuals during data collection, and the necessary ethical approval and informed consent were secured.

References

1. A. L. David and R. N. Spencer, “Clinical assessment of fetal well-being and fetal safety indicators,” *The Journal of Clinical Pharmacology,* vol. 62, no. s1, pp. S67–S78, 2022.
2. Y. Salini, S. N. Mohanty, J. V. N. Ramesh, M. Yang, and M. M. V. Chalapathi, “Cardiotocography data analysis for fetal health classification using machine learning models,” *IEEE Access,* vol. 12, no. 2, pp. 26005–26022, 2024.
3. A. M. Ponsiglione, C. Cosentino, G. Cesarelli, F. Amato, and M. Romano, “A comprehensive review of techniques for processing and analyzing fetal heart rate signals,” *Sensors,* vol. 21, no. 18, p. 6136, 2021.
4. R. C. Young, P. S. Marinescu, and N. S. Seligman, “Monitoring uterine contractions during labor: Current challenges and future directions,” *American Journal of Obstetrics and Gynecology,* vol. 228, no. 5, pp. S1192–S1208, 2023.
5. K. Barnova, R. Martinek, R. V. Kahankova, R. Jaros, V. Snasel, and S. Mirjalili, “Artificial intelligence and machine learning in electronic fetal monitoring,” *Archives of Computational Methods in Engineering,* vol. 31, no. 1, pp. 2557–2588, 2024.
6. V. Gude and S. Corns, “Integrated deep learning and supervised machine learning model for predictive fetal monitoring,” *Diagnostics,* vol. 12, no. 11, p. 2843, 2022.
7. G. W. Jiji, A. Rajesh, and A. Muthuraj, “Automated interpretation of fetal abnormalities over real-time sensory sonography using SVM classifier,” *Multimedia Tools and Applications,* vol. 83, no. 10, pp. 86207–86228, 2024.
8. M. Moutaib, M. Fattah, Y. Farhaoui, B. Aghoutane, and M. El Bekkali, “Fetal electrocardiogram prediction using machine learning: A random forest-based approach,” *Indonesian Journal of Electrical Engineering and Computer Science,* vol. 33, no. 2, pp. 1076–1083, 2024.
9. N. Innab, S. Alsubai, E. A. Alabdulqader, A. A. Alarfaj, M. Umer, S. Trelova, and I. Ashraf, “Automated approach for fetal and maternal health management using light gradient boosting model with SHAP explainable AI,” *Frontiers in Public Health,* vol. 12, no. 12, pp. 1–16, 2024.
10. Z. Cao, G. Wang, L. Xu, C. Li, Y. Hao, Q. Chen, X. Li, G. Liu, and H. Wei, “Intelligent antepartum fetal monitoring via deep learning and fusion of cardiotocographic signals and clinical data,” *Health Information Science and Systems,* vol. 11, no. 3, p. 16, 2023.
11. R. Ramya and K. Srinivasan, “Classification of amniotic fluid level using Bi-LSTM with homomorphic filter and contrast enhancement techniques,” *Wireless Personal Communications,* vol. 124, no. 12, pp. 1123–1150, 2022.
12. M. J. Khan, M. Vatish, and G. D. Jones, “PatchCTG: A patch cardiotocography transformer for antepartum fetal health monitoring,” *Sensors,* vol. 25, no. 9, p. 2650, 2025.
13. H. Liang and Y. Lu, “A CNN-RNN unified framework for intrapartum cardiotocograph classification,” *Computer Methods and Programs in Biomedicine,* vol. 229, no. 2, p. 107300, 2023.

14. M. Liu, Y. Lu, S. Long, J. Bai, and W. Lian, "An attention-based CNN-BiLSTM hybrid neural network enhanced with features of discrete wavelet transformation for fetal acidosis classification," *Expert Systems with Applications*, vol. 186, no. 12, p. 115714, 2021.
15. E. Fotiadou, R. J. G. van Sloun, J. O. E. H. van Laar, and R. Vullings, "A dilated inception CNN-LSTM network for fetal heart rate estimation," *Physiological Measurement*, vol. 42, no. 4, p. 45007, 2021.
16. G. Keerthi and M. S. Abirami, "Intelligent diagnosis of fetal organs' abnormal growth in ultrasound images using an ensemble CNN-TLFEM model," *Multimedia Tools and Applications*, vol. 83, no. 3, pp. 81167–81178, 2024.
17. P. P. A. Mohan, V. Uma, R. Sasirekha, and V. Hamsika, "FHR signal analysis using attention-based 1DCNN-BiLSTM neural network for intrapartum fetal monitoring," *Digital Signal Processing*, vol. 164, no. 9, p. 105259, 2025.
18. Z. Zhou, Z. Zhao, X. Zhang, X. Zhang, P. Jiao, and X. Ye, "Identifying fetal status with fetal heart rate: Deep learning approach based on long convolution," *Computers in Biology and Medicine*, vol. 159, no. 6, p. 106970, 2023.
19. M. Turkan, E. Dandil, F. E. Urfali, and M. Korkmaz, "FetalMovNet: A novel deep learning model based on attention mechanism for fetal movement classification in US," *IEEE Access*, vol. 13, no. 3, pp. 52508–52527, 2025.
20. A. Mehbodniya, A. J. P. Lazar, J. Webber, D. K. Sharma, S. Jayagopalan, K. Kousalya, P. Singh, R. Rajan, S. Pandya, and S. Sengan, "Fetal health classification from cardiotocographic data using machine learning," *Expert Systems*, vol. 39, no. 6, p. e12899, 2022.
21. S. Boudet, A. Houzé de l'Aulnoit, L. Peyrodie, R. Demailly, and D. Houzé de l'Aulnoit, "Use of deep learning to detect the maternal heart rate and false signals on fetal heart rate recordings," *Biosensors*, vol. 12, no. 9, p. 691, 2022.
22. N. M. Hussain, A. U. Rehman, M. T. B. Othman, J. Zafar, H. Zafar, and H. Hamam, "Accessing artificial intelligence for fetus health status using hybrid deep learning algorithm (AlexNet-SVM) on cardiotocographic data," *Sensors*, vol. 22, no. 14, p. 5103, 2022.
23. S. S. Ahmed and N. M. Mahmoud, "Early detection of fetal health status based on cardiotocography using artificial intelligence," *Neural Computing and Applications*, vol. 37, no. 6, pp. 16753–16779, 2025.
24. P. Darsana and V. N. Kumar, "Extracting fetal ECG signals through a hybrid technique utilizing two wavelet-based denoising algorithms," *IEEE Access*, vol. 11, no. 8, pp. 91696–91708, 2023.
25. Z. Yang, Y. Li, and G. Zhou, "Ts-gan: Time-series GAN for sensor-based health data augmentation," *ACM Transactions on Computing for Healthcare*, vol. 4, no. 2, pp. 1–21, 2023.
26. Y. Lin, H. Liu, L. Ruan, Z. Chen, and J. Xu, "Advancing non-invasive fetal health monitoring: A time–frequency approach to extracting fetal electrocardiogram signals," *Biomedical Signal Processing and Control*, vol. 95, no. 9, p. 106477, 2024.
27. A. J. D. Krupa, S. Dhanalakshmi, N. L. Sanjana, N. Manivannan, R. Kumar, and S. Tripathy, "Fetal heart rate estimation using fractional Fourier transform and wavelet analysis," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 4, pp. 1533–1547, 2021.
28. U. Sirisha, P. N. Srinivasu, P. Padmavathi, S. Kim, A. Pavate, J. Shafi, and M. F. Ijaz, "Dynamic multi-layer perceptron for fetal health classification using cardiotocography data," *Computers, Materials & Continua*, vol. 80, no. 2, pp. 2301–2330, 2024.