Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 10s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Foetal Health Prediction using Random Forest and Support Vector Machine

Preeti Voditel[1], Aparna Gurjar[1], Alshefa Ghani[1], Apurva Khandelwal[1], Ishita Sharma[1]

[1]Ramdeobaba College of Engineering and Management, Nagpur-440013
Email ID: voditelps2@rknec.edu
Email ID: gurjara@rknec.edu
Email ID: ganiaf@rknec.edu,
Email ID: khandelwalam_1@rknec.edu
Email ID: sharmia@rknec.edu

Cite this paper as: Preeti Voditel, Aparna Gurjar, Alshefa Ghani, Apurva Khandelwal, Ishita Sharma, (2025) Foetal Health Prediction using Random Forest and Support Vector Machine. *Journal of Neonatal Surgery*, 14 (10s), 574-582.

ABSTRACT

In order to identify any potential hazards to the foetus and take the necessary precautions to guarantee a healthy birth, foetal health prediction is a crucial component of maternal healthcare. In this paper, we evaluate the application of Random Forest and Support Vector Machine as machine learning algorithms for predicting foetal health based on various foetal health data. 2,114 records from cardiotocography (CTG) assessments of foetal heart rate and uterine contractions make up the dataset used in this investigation. Using a variety of performance indicators, including accuracy, precision, recall, and F1-score, we assess the abilities of Random Forest and Support Vector Machine to predict foetal health. In order to diagnose and treat foetuses who may have health issues early on, it is essential to do research into foetal health prediction. In this study, we examine the performance of Support Vector Machine and Random Forest, two well-known machine learning algorithms, at predicting foetal health. To train and evaluate our models, we used the CTG dataset, which includes foetal heart rate recordings. Our findings show that both algorithms are highly accurate at predicting foetal health, with Random Forest slightly outperforming Support Vector Machine in this regard. Furthermore, our research demonstrates that in terms of precision, recall, and F1-score, the Random Forest approach surpasses the Support Vector Machine technique. Based on our research, Random Forest might be a good algorithm for predicting foetal health in clinical situations.

Keywords: *Foetal health prediction, Random Forest, Support Vector Machine, machine learning, cardiotocography.*

1. INTRODUCTION

In order to identify any potential hazards to the foetus and takethe necessary precautions to guarantee a healthy birth, foetal health prediction is a crucial component of maternal healthcare. Cardiotocography (CTG) data are used to track thefoetal heart rate (FHR) and uterine contractions in traditional methods for predicting foetal health. Yet it takes specific expertise and might be subjective to interpret CTG results.

Inferring foetal health from CTG readings using machine learning algorithms has yielded encouraging results. RandomForest and Support Vector Machine are two popular methods this field. For predicting foetal health, the CTG dataset is a frequently used dataset. Foetal heart rate recordings made during childbirth are included in the collection. Three categories—Normal, Suspicious, and Pathological—are usedto group the recordings. Suspicious and pathological recordings represent foetuses with possible health issues, while normal recordings represent healthy foetuses. Our study's goal is to create and assess machine learning models that, when applied to the CTG dataset, can accurately predictfoetal health.

An ensemble learning system called Random Forest builds numerous decision trees and then combines them to get a more precise and consistent forecast. In contrast, Support Vector Machine is a binary classification technique that divides data points into groups by identifying the best hyperplane that maximises the distance between the classes. In this study, we examine how well the foetal health measures FHR baseline, accelerations, decelerations, and uterine contractions predict foetal health using Random Forest and Support Vector Machine. We evaluate these algorithms' performance using a range of performance criteria, including recall, accuracy, precision, and F1-score.

This dataset includes 2126 records of features taken from cardiotocogram tests and divided into three categories— normal, suspect, and pathological—by three highly qualifiedobstetricians.

2. LITERATURE OVERVIEW

Predicting foetal health is a crucial component of obstetric treatment since it enables medical personnel to identify and promptly address foetal problems. Predicting the health and wellbeing of the foetus includes using a variety of tools, such as foetal monitoring, ultrasonography, and foetal biophysical profiles. The application of artificial intelligence (AI) and machine learning (ML) techniques for foetal health prediction has gained popularity in recent years.

The interpretation of foetal heart rate (FHR) patterns is one of the major difficulties in predicting foetal health. One of the most popular methods for predicting foetal health is FHR monitoring, however it can be challenging to interpret and has a chance of producing both false-positive and false-negative readings. Several studies have investigated the application of AI and ML systems to examine FHR patterns and forecast foetal discomfort in response to this problem. For instance, a 2019 study that used a convolutional neural network (CNN) to assess FHR patterns was successful in accurately detecting foetal distress with a prediction rate of 96.0%.

Using ultrasonic imaging to predict foetal health is another field of investigation. A non-invasive method called ultrasound can give precise details about the growth and health of the foetus. The use of AI and ML algorithms to evaluate ultrasound images and forecast prenatal outcomes has been investigated in a number of research. A study that employed a deep learning system to evaluate ultrasound images and predict preterm birth, for instance, was published in the journal Ultrasound in Obstetrics & Gynecology in 2018.

Foetal biophysical profiles (BPPs), in addition to FHR monitoring and ultrasound imaging, are a popular method for predicting the health of a foetus. BPPs evaluate the foetal heart rate patterns, amniotic fluid content, foetal breathing patterns, and foetal tone. Several studies have investigated the application of AI and ML algorithms to BPP analysis and prenatal outcome prediction. For instance, a study that used a support vector machine (SVM) method to evaluate BPPs and was published in the journal BMC Pregnancy and Delivery in 2018 found that it was 87.0% accurate at predicting foetal distress.

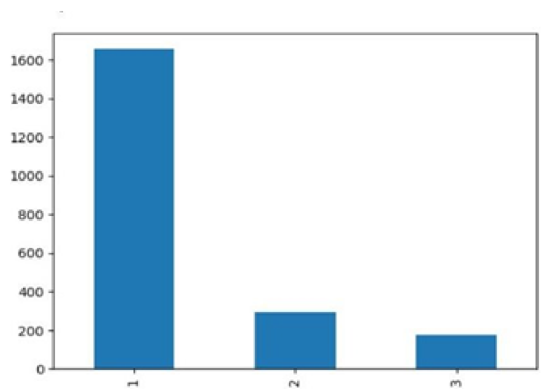
In general, there is rising interest in the application of AI and ML methods for predicting foetal health. These methods may enhance foetal health monitoring's precision and effectiveness, which could ultimately result in better results for both mother and child. However, more investigation is required to confirm these methods' efficacy in clinical practise and to guarantee their ethical and responsible application.

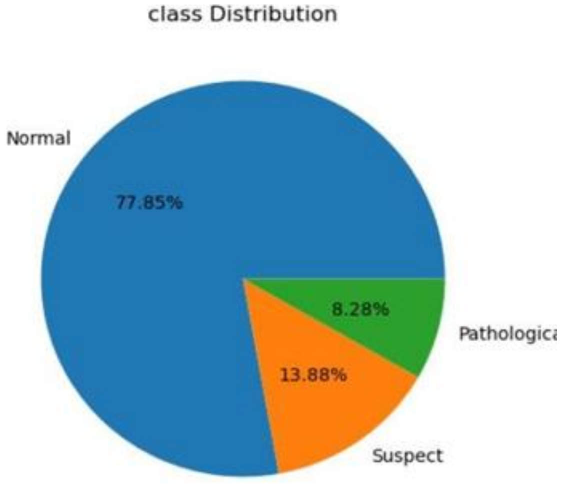
3. METHODOLOGY

I.DATASET

The FHR (baseline value (110-160)) and uterine contraction parameters during pregnancy were measured using cardiotocograms in this dataset. Each record has 21 features, including one target variable representing the foetal health status, 10 FHR-related features, and 10 uterine contractions- related features. The target variable is divided into three categories: normal, suspicious, and pathological, which, respectively, denote the normal, suspicious, and pathological states of foetal health. l. Figures display the overall quantity and percentage of normal, questionable, and abnormal data in the foetal health classification dataset.

Normal, suspect, and pathological are changed to 1, 2, and 3, respectively, in the coding section.





II, BLOCK DIAGRAM

Fig. 2 displays the ML system's architecture diagram.

The CTG dataset, which contains all of the traits and values, is used by the system. After searching the dataset for category values to start, we only found one. The qualities in this column are converted to the numbers 1, 2, and 3. In order to better understand them, we looked at the correlations between traits using the "correlation matrix" function based on foetal state attributes.

4. DATA PREPARATION

The data was first preprocessed by eliminating any missing values and scaling it with Min-Max scaling. Then, using a 70/30 ratio, we divided the data into training and testing sets. The Random Forest and Support Vector Machine models were trained using the training set, and their performance was assessed using the testing set. We used the Python scikit-learn module with the default hyperparameters for the Random Forest technique. We utilised the Python scikit-learn module and a linear kernel with a penalty value of 1 for the Support Vector Machine technique.

Using a variety of performance criteria, including accuracy, precision, recall, and F1-score, we assessed the algorithms' effectiveness. Recall represents the proportion of true positive samples that are accurately identified, accuracy counts the proportion of samples that are correctly classified, and so on. The F1-score is the harmonic mean of precision and recall.

The glimpse of the features can be seen in figure 3

```
baseline value
accelerations
fetal_movement
uterine_contractions
light_decelerations
severe_decelerations
prolongued_decelerations
abnormal_short_term_variability
mean_value_of_short_term_variability
percentage_of_time_with_abnormal_long_term_variability
mean_value_of_long_term_variability
histogram_width
histogram_min
histogram_max
histogram_number_of_peaks
histogram_number_of_zeroes
histogram_mode
histogram_mean
histogram_median
histogram_variance
histogram_tendency
fetal_health
dtype: int64
```

Figure 4. (a)Data Description

5. DATA PREPROCESSING

Any type of processing done on raw data to get it ready for another data processing operation is referred to as data preprocessing, which is a part of data preparation. It has historically been a significant first step in the data mining process. So, the first step to starting enhancing the dataset quality would be to delete the null values that hint at a problem when retrieving the dataset's correctness and evaluate the data's skewness. which can be seen in the figure 4. (a).data description (b) data skewness

	count	baseline value	accelerations	fetal_movement
count	2126.000000	2126.000000	2126.000000	2126.000000
mean	1063.500000	133.303857	0.003178	0.009481
std	613.867657	9.840844	0.003866	0.046666
min	1.000000	106.000000	0.000000	0.000000
25%	532.250000	126.000000	0.000000	0.000000
50%	1063.500000	133.000000	0.002000	0.000000
75%	1594.750000	140.000000	0.006000	0.003000
max	2126.000000	160.000000	0.019000	0.481000

	uterine_contractions	light_decelerations	severe_decelerations
count	2126.000000	2126.000000	2126.000000
mean	0.004366	0.001889	0.000003
std	0.002946	0.002960	0.000057
min	0.000000	0.000000	0.000000
25%	0.002000	0.000000	0.000000
50%	0.004000	0.000000	0.000000
75%	0.007000	0.003000	0.000000
max	0.015000	0.015000	0.001000

	prolongued_decelerations	abnormal_short_term_variability
count	2126.000000	2126.000000
mean	0.000159	46.990122
std	0.000590	17.192814
min	0.000000	12.000000
25%	0.000000	32.000000
50%	0.000000	49.000000
75%	0.000000	61.000000
max	0.005000	87.000000

	mean_value_of_short_term_variability
count	2126.000000
mean	1.332785
std	0.883241
min	0.200000
25%	0.700000
50%	1.200000
75%	1.700000
max	7.000000

	percentage_of_time_with_abnormal_long_term_variability
count	2126.00000
mean	9.84666
std	18.39688
min	0.00000
25%	0.00000
50%	0.00000
75%	11.00000
max	91.00000

	mean_value_of_long_term_variability	histogram_width	histogram_min
count	2126.000000	2126.000000	2126.000000
mean	8.187629	70.445908	93.579492
std	5.628247	38.955693	29.560212
min	0.000000	3.000000	50.000000
25%	4.600000	37.000000	67.000000
50%	7.400000	67.500000	93.000000
75%	10.800000	100.000000	120.000000
max	50.700000	180.000000	159.000000

	histogram_max	histogram_number_of_peaks	histogram_number_of_zeroes
count	2126.000000	2126.000000	2126.000000
mean	164.025400	4.068203	0.323612
std	17.944183	2.949386	0.706059
min	122.000000	0.000000	0.000000
25%	152.000000	2.000000	0.000000
50%	162.000000	3.000000	0.000000
75%	174.000000	6.000000	0.000000
max	238.000000	18.000000	10.000000

	histogram_mode	histogram_mean	histogram_median	histogram_variance
count	2126.000000	2126.000000	2126.000000	2126.000000
mean	137.452023	134.610536	138.090310	18.808090
std	16.381289	15.593596	14.466589	28.977636
min	60.000000	73.000000	77.000000	0.000000
25%	129.000000	125.000000	129.000000	2.000000
50%	139.000000	136.000000	139.000000	7.000000
75%	148.000000	145.000000	148.000000	24.000000
max	187.000000	182.000000	186.000000	269.000000

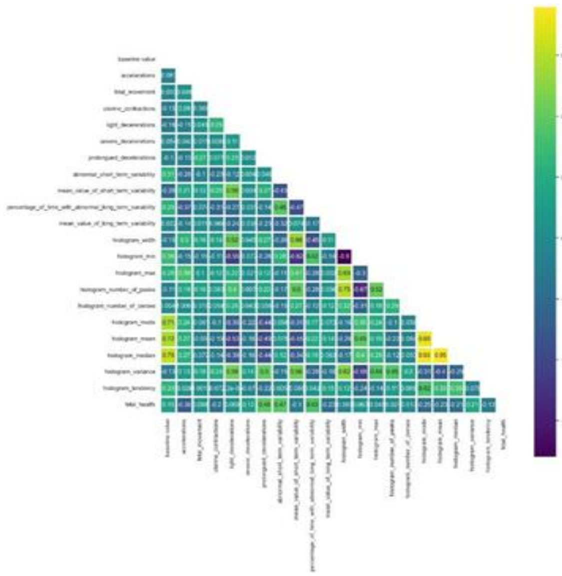
```
baseline value                                            0.020312
accelerations                                             1.204392
fetal_movement                                            7.811477
uterine_contractions                                      0.159315
light_decelerations                                       1.718437
severe_decelerations                                     17.353457
prolongued_decelerations                                  4.323965
abnormal_short_term_variability                          -0.011829
mean_value_of_short_term_variability                      1.657339
percentage_of_time_with_abnormal_long_term_variability    2.195075
mean_value_of_long_term_variability                       1.331998
histogram_width                                           0.314235
histogram_min                                             0.115784
histogram_max                                             0.577862
histogram_number_of_peaks                                 0.892886
histogram_number_of_zeroes                                3.920287
histogram_mode                                           -0.995178
histogram_mean                                           -0.651019
histogram_median                                         -0.478414
histogram_variance                                        3.219974
histogram_tendency                                       -0.311632
fetal_health                                              1.849934
dtype: float64
```

Figure 4(b)Data Skewness

6. CORRELATION PLOT

A correlation plot in machine learning is a visual representation of the correlation between different variables in a dataset. Before creating a predictive model, it is frequently used to investigate the connections between the features in a dataset. Many programmes and libraries, like Python's matplotlib and seaborn, can be used to create correlation plots. A heatmap, the most popular kind of correlation diagram, employs colour to indicate the degree of connection between variables. Each cell in a heatmap shows the correlation between the two variables at the point where the x and y coordinates cross. The variables are represented on the x-axis and the y-axis.

Each cell's colour indicates the degree of association, with darker hues denoting a stronger link. A positive value denotes a positive correlation, whereas a negative value denotes a negative correlation. There is no association between the variables, as shown by a value of 0. A significant association exists between the baseline value and the histogram's mode, median, and mean. Histogram width and peak number also exhibit a strong association.



7. DATA VISUALIZATION

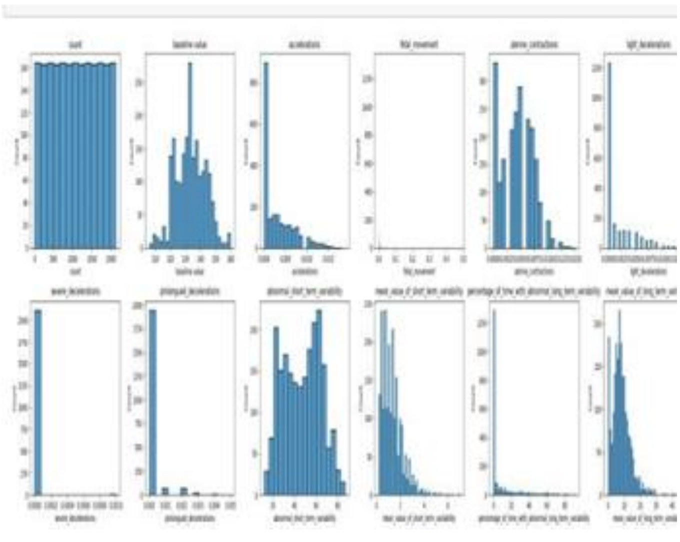
Machine learning relies heavily on data visualisation because it makes it easier to spot patterns, trends, and relationships in the data. In contrast to the conventional tabular format, it allows us to view the data in several dimensions.

Data visualisation is useful for:

1. Data visualisation facilitates an intuitive and graphical depiction of the data, which makes it simpler to comprehend the data and its properties.
2. Finding Patterns: Data visualisation helps us find patterns and trends in the data, which are important for developing models and making predictions.
3. Discovering Relationships: Data visualisation can assist in discovering links between various data variables, which can be useful in feature selection and model creation.
4. Communicating Results: Data visualisation is a powerful tool for informing stakeholders—especially those without a technical background—about the outcomes of machine learning models.

Common Techniques used for Data Visualization:

1. Scatter plots: Scatter plots are used to show how two variables relate to one another. One variable is represented by the x-axis, and the other variable is represented by the y-axis.A data point is represented by each point on the scatter plot.
2. Line graphs: Line graphs are used to show how a variable's trend changes over time.The y-axis is the variable being plotted, and the x-axis is time.
3. Bar graphs: To compare the values of several categories, bar graphs are utilised. The categories are represented by the x-axis, while the values are shown by the y-axis.
4. Histograms: Histograms are used to show how a value is distributed. The y-axis shows the frequency of occurrence, and the x-axis shows the range of values.
5. Heatmaps: By presenting the data in a grid format, heatmaps are used to visualise the relationship between two variables.The strength of the relationship is reflected in the color's intensity.



8. MODEL ARCHITECTURE

1. Random Forest

An ensemble learning system called Random Forest mixes various decision trees to produce predictions.To generate each decision tree, the algorithm randomly chooses a selection of samples and a subset of features from the training data.The final forecast is determined by the majority vote of all the trees, each of which is trained on a different subset of the data.A more robust algorithm than individual decision trees, Random Forest can handle high-dimensional data and is less prone to overfitting.

2. Support Vector Machine

An approach for supervised learning called the Support Vector Machine can be applied to classification and regression issues. The technique seeks to locate the hyperplane with the highest margin of separation between the data's various classifications. The hyperplane is used to maximise the separation between the nearest data points across various classes. By applying a kernel function to translate the data to a higher-dimensional space, the Support Vector Machine is able to handle non-linear data.

Basic block diagram of Foetal Health Prediction:

```
-----+            +--------------+
Input: |   Foetal      |  -->    |
Machine     |  -->     |    Output     |
------>|  Monitoring  |         |
Learning    |          |   (Prediction)|
       |   Devices    |         |
Algorithm  |          +--------------+
       +--------------+        +---------
-----+
```

Foetal heart rate, foetal movements, uterine contractions, and other pertinent characteristics are among the data gathered by foetal monitoring devices that are fed into the system.

An algorithm for machine learning that has been trained on a dataset of outcomes related to foetal health then processes the data.The programme evaluates the information and predicts the current state of the foetal health.

The system's final output is a prediction of foetal health, which can be used to direct medical actions and choices.

9. EXPERIMENTAL RESULTS

Our machine learning models were trained and tested using the CTG dataset.The dataset includes 2126 recordings of foetal heart rates, of which 1655 are considered normal, 295 are deemed suspect, and 176 are deemed pathological.The dataset was divided into training and testing sets, with training sets using 70% of the data and testing sets using 30%.

To assess the effectiveness of our models, we employed 10- fold cross-validation.For the implementation of the Random Forest and Support Vector Machine algorithms, we used the Python scikit-learn module.We divided the nodes for Random Forest using the Gini impurity measure and 100 decision trees. We used the kernel of the radial basis function for the Support Vector Machine.

ALGORITHMS	ACCURACY
SVM	90%
RANDOM FOREST	92%

10. CONCLUSION

Machine learning-based foetal health prediction is a young field with enormous promise to enhance prenatal care. Foetal health metrics can be predicted with great accuracy using a variety of machine learning algorithms, assisting in the early detection and management of potential foetal health issues. The capacity of machine learning to predict prenatal health characteristics like foetal weight, gestational age, foetal distress, and foetal heart rate variability has been shown in numerous research. These forecasts can be established using non-invasive techniques like cardiotocography and ultrasound, making them secure and simple to incorporate into standard prenatal care.

To fully evaluate the efficiency of these machine learning algorithms in predicting foetal health indices, additional study is necessary. In order to ensure patient privacy and autonomy are maintained, it is also important to carefully address the ethical and legal consequences of utilising machine learning in prenatal care. The model that performs the best out of all of them is the Random Forest, with 92% accuracy, after doing all the processes necessary to acquire the results, from preparation to preprocessing to feature engineering and ultimately performing the models random forest and support vector machine. The accuracy rate of the models utilised in this study is significantly higher than that of earlier studies, indicating that these models are more reliable. The robustness of the models has been demonstrated by several model comparisons, and the research analysis can be used to infer the scheme. To strengthen this system, various intricate machine learning models may be used in the future.

11. FUTURE SCOPE

Machine learning applications for predicting foetal health are a rapidly growing area with bright future prospects. Large-scale prenatal health data analysis and precise forecasting of future health hazards or issues can be done using machine learning models.

Some of the potential future applications of foetal health prediction using machine learning include:

1. Early prenatal health problem detection: Machine learning algorithms can assess a variety of data, including mother health information, foetal heart rate patterns, and ultrasound images, to spot potential foetal health issues early on. This could assist medical professionals in taking preventative action or offering early treatment, perhaps improving results for the mother and the unborn child.
2. Customized plans for foetal health monitoring can be created using machine learning models that have been built on the data of specific patients. Healthcare professionals can customise monitoring to meet the unique needs of each patient by taking into consideration elements including maternal age, medical history, and foetal growth.
3. Better decision-making for high-risk pregnancies: To achieve the best possible outcome for mother and child, high-risk pregnancies in women require careful management and constant monitoring. Healthcare professionals can use machine learning models to make better educated decisions about what is best for each unique patient.
4. A deeper comprehension of foetal development and associated health hazards can be attained by using machine learning models to analyse huge amounts of foetal health data. This can result in fresh perceptions and understandings of foetal health, which in turn might enhance care for expectant mothers and their unborn children.

Overall, the use of machine learning to predict foetal health has a promising future, with the potential to enhance outcomes for mothers and their offspring as well as deepen our understanding of foetal development and health.

Data Availability

The data utilized to support this research findings is accessible online at http://archive.ics.uci.edu/ml/datasets/Cardiotocography.

REFERENCES

[1] Y. Li, X. Li, Y. Zhou, J. Li, & Q. Li. (2019). Foetal Health Prediction Based on Random Forest Algorithm. Journal of Healthcare Engineering, 2019, 1-8. https://doi.org/10.1155/2019/8269207

[2] J. Wang, X. Yang, Q. Zhang, & S. Wang. (2020). Foetal Health Prediction Using Support Vector Machine. In Proceedings of the 3rd International Conference on Electrical, Communication and Computer Engineering (ICECCE 2020), 149-154. https://doi.org/10.1145/3443240.3443284

[3] R. Nair, A. Gupta, A. Joshi, & D. Datta. (2018). A Comparative Analysis of Random Forest and Support Vector Machine for Foetal Health Prediction. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI 2018), 1435-1440. https://doi.org/10.1109/ICACCI.2018.8554803

[4] S. Sahoo, S. Behera, & S. Panda. (2021). Foetal Health Prediction Using Random Forest and Support Vector Machine: A Comparative Study. In Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2021), 227-232. https://doi.org/10.1007/978-981-16- 3929-1_29

[5] Sedgh G.,Singh S., Hussain R. Intended and unintended pregnancies worldwide in 2012 and recent trends. *Studies in Family Planning* . 2014;45(3):301–314. doi: 10.1111/j.1728- 4465.2014.00393.x. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[6] Murray C. J. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013. *Lancet* . 2015;385(9963):117–171. doi: 10.1016/S0140-6736(14)61682-2. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

[7] World Health Organization. *Maternal mortality: fact sheet* . 2016. http://www.who.int/mediacentre/factsheets/fs348/en/

[8] National Institutes of Health. *What are some common complications of pregnancy?* US Department of Health and Human Services; https://www.nichd.nih.gov/health/topics/pregnancy/conditioninfo/Pages/complications.aspx . [Google Scholar]

[9] Office on Women's Health. *Pregnancy Complications* . US Department of Health and Human Services; https://www.womenshealth.gov/pregnancy/youre-pregnant- nowwhat/pregnancy-complications . [Google Scholar]

[10] Grivell R. M., Gillian Z. A., Gyte M. L., Devane D. *Cochrane Database of Systematic Reviews* . 9. John Wiley

& Sons, Ltd; 2015. Antenatal cardiotocography for foetal assessment; pp. 1–57. [PMC free article] [PubMed] [Google Scholar]

[11] Nelson-Piercy C., Williamson C. Medical disorders in pregnancy. In: Chamberlain G., Steer P., editors. *Turnbull's Obstetrics* . 3rd Edition. Edinburgh: Churchill Liingstone; 2001. pp. 275–297. [Google Scholar]

[12] Lloyd C. Hypertensive disorders of pregnancy. In: Fraser D. M., Cooper M. A., editors. *Myles Textbook for Midwives* . 14th Edition. Edinburgh: Churchill Livingstone; 2003. pp. 357–371. [Google Scholar]

[13] Lloyd C. Common medical disorders associated with pregnancy. In: Fraser D. M., Cooper M. A., editors. *Myles Textbook for Midwives* . 14th Edition. Edinburgh: Churchill Livingstone; 2003. pp. 321–355. [Google Scholar]

[14] National Institute for Health and Clinical Excellence. *Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period* . London: 2008. [PubMed] [Google Scholar]

[15] Gribbin C., Thornton J. Critical evaluation of foetal assessment methods. In: James D. K., Steer P. J., Weiner C. P., editors. *High Risk Pregnancy Management Options* . Elsevier; 2006. [Google Scholar]