A CLINICAL DECISION SUPPORT SYSTEM FOR DIAGNOSING AND PREVENTING MATERNAL DISEASES

GERALD GAITOS

A CLINICAL DECISION SUPPORT SYSTEM FOR DIAGNOSING AND PREVENTING MATERNAL DISEASES

INTRODUCTION

Due to the declining maternal health and an increasing number of maternal disease complications and deaths, the project's primary purpose is to develop and design a machine learning model using the Maternal Health Risk Dataset by Ahmed et al. (2020) and to serve as an expert system to help identify women at risk of maternal diseases and to prevent maternal disease complications from happening.

This project intends to provide information about the importance of monitoring vital signs benefiting women, healthcare providers, and researchers in identifying maternal disease risks and complications and preventing women at risk from acquiring them. This project also presents the Maternal Health Risk Dataset, the materials and methods, the algorithms, the result and discussion, and the references.

BACKGROUND

In most parts of the world, maternal health has declined since 2016. No progress has been noted for the past 15 years (Taylor, 2023). In 2014, maternal mortality was excessively high, with about 800 women dying from preventable causes. They died from preventable pregnancy- and childbirth-related causes and complications (WHO, 2014). According to the World Health Organization (2023), maternal health is not limited to pregnancy and childbirth but includes the postnatal period. Most of these women experienced complications that developed following pregnancy and childbirth. These significant complications are the following: severe bleeding (after childbirth), infections (after childbirth), high blood pressure during pregnancy (pre-eclampsia and eclampsia), and unsafe abortion (WHO, 2014).

High blood pressure during pregnancy, including preeclampsia/eclampsia, chronic hypertension, gestational hypertension, and preeclampsia superimposed on chronic hypertension, is considered the second most common cause of maternal mortality (Ansari et al., 2019). Disorders related to abnormal blood pressure are preventable with accurate blood pressure monitoring, helping reproductive and pregnant women prevent the debilitating health consequences (Garovic et al., 2021). Mugyenyi et al. (2021) state that vital sign monitoring is crucial in identifying maternal complications and intervening when needed.

This project focuses on developing and designing a machine learning model using the Maternal Health Risk Dataset (Ahmed et al., 2020) that will help serve as an expert system to identify and stratify women at risk of maternal diseases and to prevent them from having complications.

PROJECT OBJECTIVES

The main objective of this project was to develop and design a machine learning model that could be utilized as an expert system or a clinical decision support system that would help medical professionals to determine the risk of women having maternal diseases and possibly prevent their debilitating complications. This project allowed women to be treated timely and effectively, limiting the cases of mortality and morbidity from these preventable diseases.

A CLINICAL DECISION SUPPORT SYSTEM FOR DIAGNOSING AND PREVENTING MATERNAL DISEASES

This goal was achieved by utilizing the Maternal Health Risk Dataset (Ahmed et al., 2020) to build a model that could be used to diagnose and even prevent cases of maternal diseases and test the model for accuracy and reliability.

SPECIFIC AIMS

The project focused on building a machine learning model and clinical decision support system for medical professionals to determine maternal disease risks and prevent complications. To achieve a machine learning model and clinical decision support system, the project uses the Maternal Health Risk Dataset (Ahmed et al., 2020) to train and test a model that can accurately identify women at risk of having maternal diseases. This project has three aims:

Primary: To determine which attributes significantly contribute to high and low maternal health risks.

Secondary: To build training and testing models from Maternal Health Risk Dataset using different classification algorithms.

Tertiary: To develop and design a machine learning model that can serve as an expert system with the highest accuracy and reliability.

MATERIALS AND METHODS

The project applied Python and R programming to the Maternal Health Risk Dataset in developing and designing a machine learning model and clinical decision support system that determined the maternal health risks of women. In addition, the dataset underwent data preprocessing to eliminate noise, inconsistency, and missing data values and improve the results' accuracy and reliability.

For the primary aim, the Maternal Health Risk Dataset was subjected to statistical tools using R programming to determine the significance of each attribute concerning maternal health risks.

For the secondary and tertiary aims, the project used four classification methods, namely, K nearest neighbors, decision tree, Naïve Bayes, logistic regression, and support vector machine, to develop and design the training and testing models and use the model with the highest accuracy and reliability for clinical decision support system development using Python.

DATA/OTHER RESOURCES NEEDED

This project utilized the Maternal Health Risk Dataset (Ahmed et al., 2020), which includes information from different hospitals, community clinics, and maternal health care in Bangladesh. The dataset contains seven attributes, namely, Age, SystolicBP, DiastolicBP, BS, BodyTemp, HeartRate, and RiskLevel, with 1,014 instances. It is in a CSV format and can be accessed at http://archive.ics.uci.edu/ml/datasets/Maternal+Health+Risk+Data+Set#.

A CLINICAL DECISION SUPPORT SYSTEM FOR DIAGNOSING AND PREVENTING MATERNAL DISEASES

EXPERIMENTAL SETUP

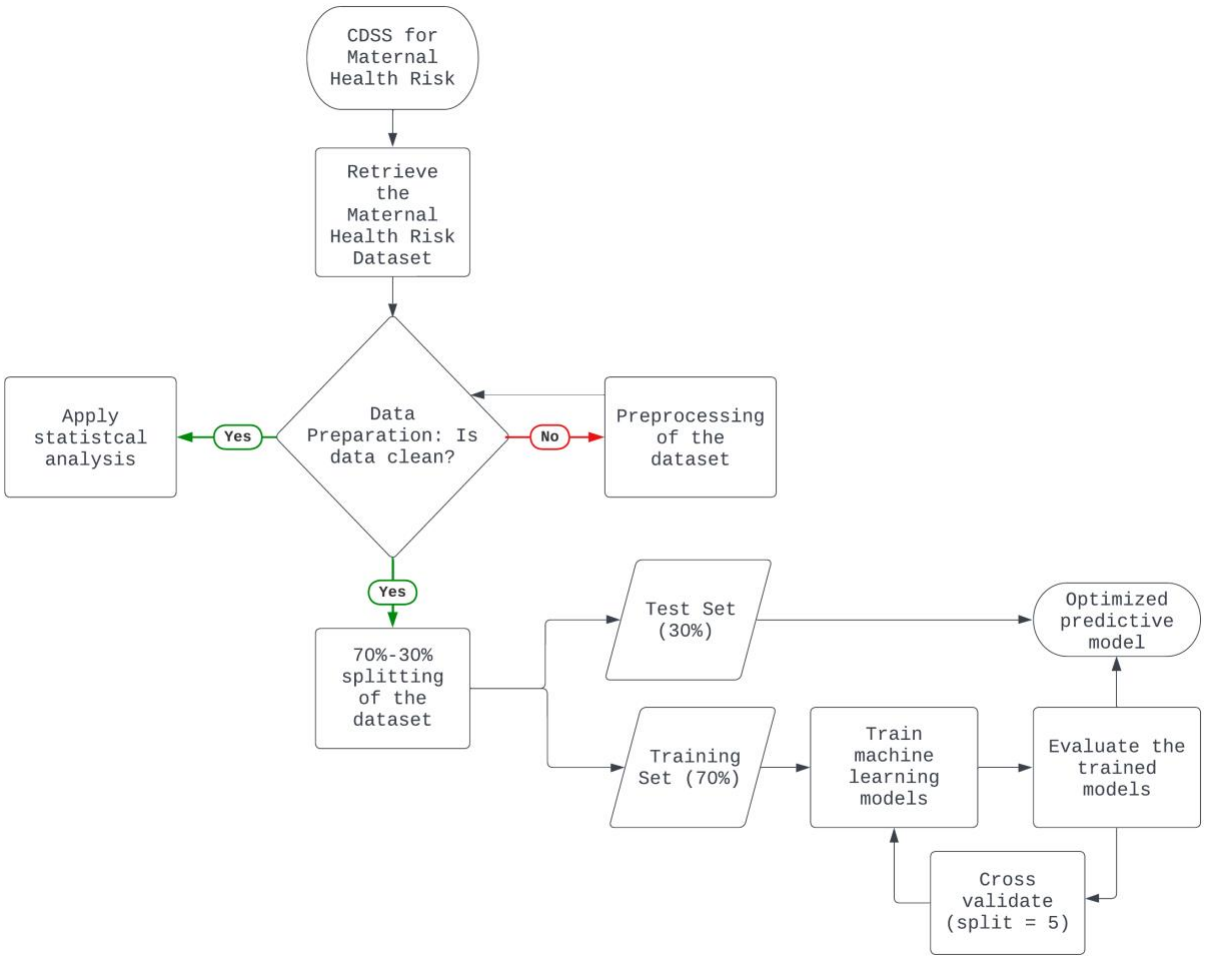


Figure 1. The flowchart of the machine learning process

Figure 1 shows the general steps taken to obtain the optimized predictive model utilized and incorporated for creating and designing the expert system. The dataset was obtained and subjected to data preparation to ensure no noise, inconsistency, or missing data values. Statistical analysis was applied to determine the association among variables. The dataset was split into 70% training and 30% test sets. The training set was subjected to five machine learning algorithms, evaluated, and cross-validated in five folds. The optimized predictive model was obtained and incorporated to create and design the expert system.

PREPROCESSING

The Maternal Health Risk Dataset underwent data cleaning to ensure the data was free from missing values, outliers, and inconsistencies. Figure 2 shows no missing values in the dataset, the first five

rows of the dataset, and the data types of the variables. Statistical analysis was applied to the preprocessed dataset to determine the association of the six attributes (Age, SystolicBP, DiastolicBP, BS, BodyTemp, and HeartRate) to the outcome attribute RiskLevel.

Initially, the dataset contained multi-class outcomes for the attribute RiskLevel. The attribute RiskLevel was converted into binary outcomes to have a more balanced class distribution. The dataset was divided into 70% training and 30% test sets. The training set was subjected to five machine learning algorithms: K nearest neighbors, decision tree, Naïve Bayes, support vector machine, and logistic regression. All the created training models were evaluated and cross-validated in five folds to ensure that the models were generalizable to the unseen data. The cross-validation scores were computed. The training models were tested using the test set. The accuracy and balanced accuracy scores were computed. Finally, the best-performing machine learning model was applied to be used as an expert system.

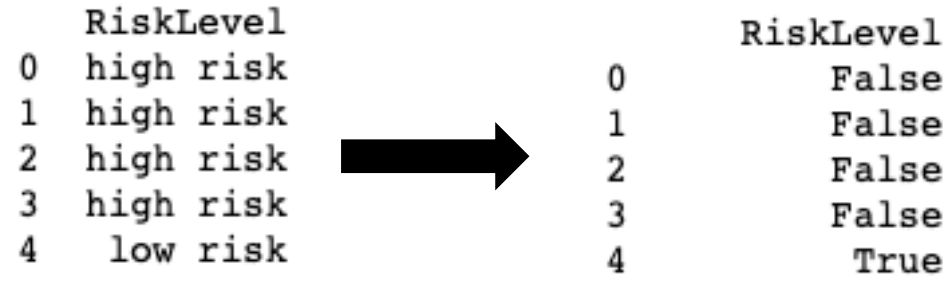
```
missing_values = dataframe.isna()

# Use the sum() method to count the number of missing values in each colum
missing_counts = missing_values.sum()

# Print the number of missing values in each column
print(missing_counts)
```

```
Age             0
SystolicBP      0
DiastolicBP     0
BS              0
BodyTemp        0
HeartRate       0
RiskLevel       0
dtype: int64
```

	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
0	25	130	80	15.0	98.0	86	high risk
1	35	140	90	13.0	98.0	70	high risk
2	29	90	70	8.0	100.0	80	high risk
3	30	140	85	7.0	98.0	70	high risk
4	35	120	60	6.1	98.0	76	low risk

```
(1014, 7)
Age               int64
SystolicBP        int64
DiastolicBP       int64
BS              float64
BodyTemp        float64
HeartRate         int64
RiskLevel        object
dtype: object
```

Figure 2.1 Preprocessing and displaying the head of the dataset.



	RiskLevel
0	high risk
1	high risk
2	high risk
3	high risk
4	low risk

	RiskLevel
0	False
1	False
2	False
3	False
4	True

Figure 2.2 Preprocessing and displaying the dataset after conversion from 3 to 2 classes.

RESULTS

DETERMINATION OF ASSOCIATION AMONG ATTRIBUTES

After data preprocessing, the signification association among attributes was determined using the general linear model in R programming, as seen in Figure 3.

```
Call:
glm(formula = RiskLevel ~ Age + SystolicBP + DiastolicBP + BS +
    BodyTemp + HeartRate, family = binomial(), data = maternal)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4133  -0.1261   0.3228   0.4827   2.9770

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 71.211782   7.684228   9.267  < 2e-16 ***
Age          0.016681   0.009974   1.672 0.094435 .
SystolicBP  -0.028424   0.009794  -2.902 0.003706 **
DiastolicBP -0.046220   0.012215  -3.784 0.000154 ***
BS          -0.496913   0.045579 -10.902  < 2e-16 ***
BodyTemp    -0.561051   0.071837  -7.810 5.72e-15 ***
HeartRate   -0.051243   0.012664  -4.046 5.20e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1179.29  on 1013  degrees of freedom
Residual deviance:  670.32  on 1007  degrees of freedom
AIC: 684.32

Number of Fisher Scoring iterations: 6
```

Figure 3. General Linear Model on the dataset.

Figure 3 shows that the attributes SystolicBP, DiastolicBP, BS, BodyTemp, and HeartRate demonstrated p-values of <0.05, suggesting that these attributes were significantly associated with the outcome, RiskLevel. Therefore, it could be interpreted that extreme values among these attributes significantly contributed to women's risk of getting maternal diseases.

BUILDING PREDICTIVE MODELS

Since the association of the attributes was established, predictive models were created from the dataset through the five machine learning algorithms: K nearest neighbors, decision tree, Naïve Bayes, support vector machine, and logistic regression. To ensure that each created model was generalizable to new and unseen data, cross-validation in five folds was done. The cross-validation, accuracy, and balanced accuracy scores were computed in Table 1.

Table 1. Predictive Models

Machine Learning Algorithm	Average Cross-validation Score (5 folds)	Accuracy Score	Balanced Accuracy Score
K=4 nearest neighbors	0.79	0.94	0.89
Decision tree	0.76	0.91	0.88
Naïve Bayes	0.74	0.56	0.66
Support vector machine (kernel=poly)	0.76	0.64	0.71
Logistic Regression	0.76	0.65	0.62

Table 1 shows that among the five machine learning models, the K=4 nearest neighbor training model demonstrated the best accuracy and balanced accuracy scores of 0.94 and 0.89, respectively. The training models demonstrated average cross-validation scores of around 0.74 to 0.79. The accuracy score of the best-performing machine learning model suggested that the model could correctly predict the outcome attribute RiskLevel, 94% of the time. The balanced accuracy score of 89% suggested that the model was not biased toward any single outcome class. Despite having high accuracy and balanced accuracy scores, the cross-validation scores suggested that the trained models could still undergo improvement to ensure the generalizability of the unseen and new data.

APPLICATION TO AN EXPERT SYSTEM

After identifying the best-performing machine learning model by determining the highest accuracy and balanced accuracy scores, the K=4 nearest neighbor predictive model created an expert system that women could utilize to stratify their risk levels. Figure 3 and 4 show the function created and the sample utilization of the designed expert system. The values inputted to the program were from the original dataset. This expert system was expected to predict approximately 94% correct risk level stratifications.

```
##### CDSS #####
def predict_risk_level_from_user_input():
    age = input("Enter age (in years): ")
    sbp = input("Enter systolic blood pressure (in mmHg): ")
    dbp = input("Enter diastolic blood pressure (in mmHg): ")
    bs = input("Enter blood sugar level (in mmol/L): ")
    temp = input("Enter body temperature (in F): ")
    hr = input("Enter heart rate (in bpm): ")

    # Create a dataframe with the user input values
    input_df = pd.DataFrame({
        'Age': [age],
        'SystolicBP': [sbp],
        'DiastolicBP': [dbp],
        'BS': [bs],
        'HeartRate': [hr],
        'BodyTemp': [temp]
    })

    # Use the pre-trained model to make a prediction
    risk_level = model.predict(input_df)

    # Return the predicted risk level
    return risk_level[0]
```

Figure 3. Function created for expert system

```
#25,130,80,15,98,86
#expected result: high risk --> AT RISK
predicted_risk_level = predict_risk_level_from_user_input()

print("Are you at RISK of maternal diseases?")
if not predicted_risk_level == True:
    print('YES, seek medical help!')
else:
    print('NO, you are in good health!')
```

```
Enter age (in years): 25
Enter systolic blood pressure (in mmHg): 130
Enter diastolic blood pressure (in mmHg): 80
Enter blood sugar level (in mmol/L): 15
Enter body temperature (in F): 98
Enter heart rate (in bpm): 86
Are you at RISK of maternal diseases?
YES, seek medical help!
```

```
#15,120,80,7.01,98,70
#expected result: low risk --> NOT AT RISK
predicted_risk_level = predict_risk_level_from_user_input()

print("Are you at RISK of maternal diseases?")
if not predicted_risk_level == True:
    print('YES, seek medical help!')
else:
    print('NO, you are in good health!')
```

```
Enter age (in years): 15
Enter systolic blood pressure (in mmHg): 120
Enter diastolic blood pressure (in mmHg): 80
Enter blood sugar level (in mmol/L): 7.01
Enter body temperature (in F): 98
Enter heart rate (in bpm): 70
Are you at RISK of maternal diseases?
NO, you are in good health!
```

```
#35,120,80,6.9,98,78
#expected result: mid risk --> AT RISK
predicted_risk_level = predict_risk_level_from_user_input()

print("Are you at RISK of maternal diseases?")
if not predicted_risk_level == True:
    print('YES, seek medical help!')
else:
    print('NO, you are in good health!')
```

```
Enter age (in years): 35
Enter systolic blood pressure (in mmHg): 120
Enter diastolic blood pressure (in mmHg): 80
Enter blood sugar level (in mmol/L): 6.9
Enter body temperature (in F): 98
Enter heart rate (in bpm): 78
Are you at RISK of maternal diseases?
YES, seek medical help!
```

```
#22,85,60,6.9,98,76
#expected result: mid risk --> AT RISK
predicted_risk_level = predict_risk_level_from_user_input()

print("Are you at RISK of maternal diseases?")
if not predicted_risk_level == True:
    print('YES, seek medical help!')
else:
    print('NO, you are in good health!')
```

```
Enter age (in years): 22
Enter systolic blood pressure (in mmHg): 85
Enter diastolic blood pressure (in mmHg): 60
Enter blood sugar level (in mmol/L): 6.9
Enter body temperature (in F): 98
Enter heart rate (in bpm): 76
Are you at RISK of maternal diseases?
YES, seek medical help!
```

```
#30,120,80,6.9,101,76
#expected result: mid risk --> AT RISK
predicted_risk_level = predict_risk_level_from_user_input()

print("Are you at RISK of maternal diseases?")
if not predicted_risk_level == True:
    print('YES, seek medical help!')
else:
    print('NO, you are in good health!')
```

```
Enter age (in years): 30
Enter systolic blood pressure (in mmHg): 120
Enter diastolic blood pressure (in mmHg): 80
Enter blood sugar level (in mmol/L): 6.9
Enter body temperature (in F): 101
Enter heart rate (in bpm): 76
Are you at RISK of maternal diseases?
NO, you are in good health!
```

```
#19,120,80,7,98,70
#expected result: mid risk --> AT RISK
predicted_risk_level = predict_risk_level_from_user_input()

print("Are you at RISK of maternal diseases?")
if not predicted_risk_level == True:
    print('YES, seek medical help!')
else:
    print('NO, you are in good health!')
```

```
Enter age (in years): 19
Enter systolic blood pressure (in mmHg): 120
Enter diastolic blood pressure (in mmHg): 80
Enter blood sugar level (in mmol/L): 7
Enter body temperature (in F): 98
Enter heart rate (in bpm): 70
Are you at RISK of maternal diseases?
NO, you are in good health!
```

Figure 4. The sample function was developed to use an expert system—sample user-input values and resulting predicted risk levels.

Figure 4 shows the code used to create and design a user-input expert system that would display and stratify the user's risk level based on the user's age, systolic blood pressure, diastolic blood pressure, blood sugar level, body temperature, and heart rate. Four out of six runs showed correct predictions upon testing the expert system using random data from the original dataset. However, incorrect predictions were observed from the mid risk level class.

DISCUSSION

This project demonstrated that among the six attributes, only five showed signification associations with the outcome attribute RiskLevel. These attributes were SystolicBP, DiastolicBP, BS, BodyTemp, and HeartRate, with p-values of <0.05. From this result, it can be inferred that discrepancies and abnormalities in the mentioned attributes increase the risk of acquiring

maternal diseases. This finding was supported by Garovic et al. (2021) and Mugyenyi et al. (2021). They expressed the importance of using and knowing basic vital signs, as these could help prevent women from getting at risk of maternal diseases and diagnose women with maternal diseases.

Moreover, this project also illustrated that among the five machine learning algorithms, K=4 nearest neighbors resulted in the highest accuracy and balanced accuracy scores of 0.94 and 0.89, respectively, which could be used as an expert system. However, Ahmed et al. (2020) utilized the same dataset using the Internet of Things. They found that Logistic Regression Model Tree gave the highest accuracy in the case of classification and prediction of the risk levels, which was opposite to what was found in this project. In addition, the cross-validation scores computed for each training model showed values around 0.74 to 0.79, suggesting that the models needed more improvement to generalize new and unseen data. Despite the contradiction in the best-performing machine learning model, the model obtained from this project was successfully utilized as an expert system.

CONCLUSION

From this analysis, the attributes SystolicBP, DiastolicBP, BS, BodyTemp, and HeartRate showed signification associations with the outcome RiskLevel relating to increased risk of acquiring maternal diseases. Furthermore, it was found through machine learning that the K=4 nearest neighbor model had the highest accuracy for the classification and prediction of the risk levels. Therefore, it could be utilized effectively as an expert system. This project would benefit women with better health outcomes and healthcare professionals to prevent the increase in maternal diseases and eliminate the decreasing trend in maternal health. For future work, using a larger dataset to improve the cross-validation, accuracy, and balanced accuracy scores would be beneficial. Incorporating other machine learning algorithms to compare and attain the best-performing machine learning model.

A CLINICAL DECISION SUPPORT SYSTEM FOR DIAGNOSING AND PREVENTING MATERNAL DISEASES

APPENDIX

DATA DICTIONARY (Ahmed et al., 2020)

Attribute	Description	Data Type	Unit Measurement
Age	Any ages in years when a women during pregnant	Integer	year
SystolicBP	Upper value of Blood Pressure in mmHg, another significant attribute during pregnancy	Integer	mmHg
DiastolicBP	Lower value of Blood Pressure in mmHg, another significant attribute during pregnancy	Integer	mmHg
BS	Blood glucose levels is in terms of a molar concentration, mmol/L	Integer	mmol/L
BodyTemp	Temperature of the body	Integer	F
HeartRate	A normal resting heart rate in beats per minute	Integer	bpm
RiskLevel	Predicted Risk Intensity Level during pregnancy considering the previous attribute	Nominal	N/A

REFERENCES

Ahmed M., Kashem M.A., Rahman M. & Khatun S. (2020). *Maternal Health Risk Dataset* [Dataset]. UCI Machine Learning Repository. Retrieved March 26, 2023, from http://archive.ics.uci.edu/ml/datasets/Maternal+Health+Risk+Data+Set#

Ahmed M., Kashem M.A., Rahman M. & Khatun S. (2020). *Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT)*. Lecture Notes in Electrical Engineering, vol 632. Springer, Singapore. https://doi.org/10.1007/978-981-15-2317-5_30

Ansari, N., Manalai, P., Maruf, F., Currie, S., Stekelenburg, J., van Roosmalen, J., Kim, Y. & Tappis, H. (2019). Quality of care in early detection and management of pre-eclampsia/eclampsia in health facilities in Afghanistan. *BMC Pregnancy and Childbirth* **19** (36). https://doi.org/10.1186/s12884-018-2143-0

Garovic, V. D. et al. (2021 December 15). Hypertension in Pregnancy: Diagnosis, Blood Pressure Goals, and Pharmacotherapy: A Scientific Statement From the American Heart Association. *Hypertension 2022,* 79, e21-e41. https://doi.org/10.1161/HYP.0000000000000208

Hunt, P. & Bueno de Mesquita, J. (2010, January 1). Reducing Maternal Mortality: The contribution of the right to the highest attainable standard of health. University of Essex. Retrieved April 3, 2023, from https://www.unfpa.org/sites/default/files/pub-pdf/reducing_mm.pdf

Mugyenyi, G. R. et al. (2021 March 11). Quality of vital sign monitoring during obstetric hospitalizations at a regional referral and teaching hospital in Uganda: an opportunity for improvement. *Pan Afr Med J 2021*, 38, 252. doi: 10.11604/pamj.2021.38.252.21749

Taylor, L. (2023). Maternal health is now deteriorating in much of the world, UN report shows. *BMJ 2023*, 380, 454. doi:10.1136/bmj.p454

World Health Organization. (2014). Maternal mortality. *World Health Organization.* Retrieved April 29, 2023, from https://apps.who.int/iris/bitstream/handle/10665/112318/WHO_RHR_14.06_eng.pdf

World Health Organization. (2023). Maternal health. *World Health Organization*. Retrieved April 29, 2023, from https://www.who.int/health-topics/maternal-health#tab=tab_1