A Bayesian Model Approach to predicting Maternal Health Risk
Fatima Essa - 2107007

***Abstract*— This report is based on an implementation of a Bayesian Probabilistic Graphical Model in the field of Maternal Health Risk. The various components of the paper will delve deeper into said implementation and execution.**

I. INTRODUCTION

The medical field we know as healthcare is an extremely dynamic and volatile area of practice. Predicting and understanding health outcomes has always been of great significant and with the strides made in the domain of Computer Science it is now possible to combine those fields.The maternal health domain play a significant role of importance upon the people of today and the generations of the future. The success of this field and its continuity is essential to human life. Significant strides have been made in the past to subsequently reduce maternal mortality and morbidity rates, challenges are still prevalent. A global concern lies in the ability to correctly predict potential health risks that a mother may face during pregnancy, childbirth, and the postpartum period.

The field of Probabilistic graphical models (PGMs) is a promising avenue to explore in order to address this concern. PGMs represent a set of random variables and their conditional dependencies through a graph. Through the various factors influencing maternal health, a complex network of relationships can be established, providing us with a deeper understanding as to how different variables interplay and contribute to the overall health risk profile of an expectant mother.

PGMs allow us to pave the way for both understanding observed data and making predictions about unseen/future data.

A. *Problem Statement*

Vast amounts of data have been collected over the years in relation to maternal health, yet there still remains a gap in utilising this data to build predictive models that can guide clinical decisions. Int his assignment we aim to address the following: How can we best leverage the power of probabilistic graphical models to build a robust, reliable, and interpretable model that predicts maternal health risks? Through the answering of this problem statement we aim to contribute to the broader goal of improving maternal health outcomes globally.

II. METHODOLOGY

The method in which the assignment was carried out in terms of data collection, data pre-processing, PGM implementations, and result analysis will be discussed in this section.

A. *Data Collection and Preprocessing*

The data used within this assignment was collected from Kaggle (https://www.kaggle.com/datasets/csafrit2/maternal-health-risk-data). The data has been collected from various hospitals, community clinics, maternal health cares through the IoT based risk monitoring system. The features of the dataset are as follows:

TABLE I
DATASET

Age	Age in years when a woman is pregnant.
SystolicBP	Upper value of Blood Pressure in mmHg.
DiastolicBP	Lower value of Blood Pressure in mmHg.
BS	Blood glucose levels (mmol/L).
HeartRate	Normal resting heart rate (bpm).
Risk Level	Predicted Risk Intensity Level during pregnancy.

The object of data pre-processing is to prepare the dataset for analysis and computations. It accounts for any anomalies and/or discrepancies by cleaning and transforming the data.

- Filtering: Any outliers in terms of HeartRate were removed to eliminate the possibility of any erroneous data.
- Missing Values: Any numerical feature with missing values are replace with the mean of the respective numerical feature. Any categorical feature with missing values are replace with the mode of the respective categorical feature.
- Standardization: All numerical features are standardized to bring them to a common scale and reduce variability.

The dataset was split into 2 sets: a training set and a validation set. This was done using StratifiedShuffleSplit as it ensures that the distribution of the RiskLevel remains consistent across both sets. This is essential when it comes to creating our model as it reduces any possible class imbalances amongst the two sets. 80% of the data was used for training, and 20% was reserved for testing/validation.

B. *Bayesian Network*

A Bayesian Network is a specific type of Probabilistic Graphical Model (PGM), whose network solely represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG). A Bayesian Network models a distribution in the form of a graph (G), where:

- G is a data structure that provides the skeleton for representing a joint distribution compactly in a factorised way.

- G is a compact set of conditional independence assumptions about the joint distribution.

Nodes represent variables and edges depict the probabilistic dependencies between them. The direction of the relationship, and the strength of said relationship, is determined by conditional probability distributions (CPDs). Given the nature of our dataset and the below generated correlation matrix,

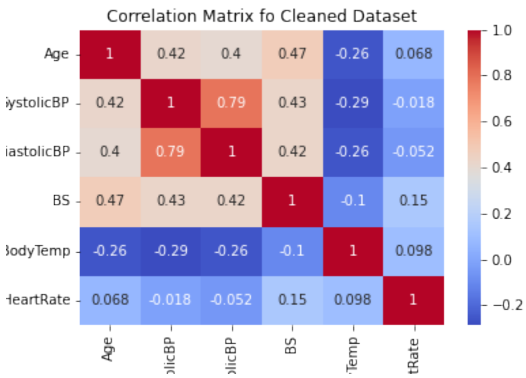

Fig. 1. Bayesian Network

it can be seen that causal relationships, prediction and incorporating prior knowledge are all aspects of what we are trying to achieve. By looking at the above matrix and the goal of this assignment, it is clear that a Bayesian Network Model is ideal.

Hill Climbing Approach: Structure determination is about figuring out which variables are dependent on which others and in what ways. To address this we have implemented the HillClimbSearch. It is a heuristic search technique. The search begins with with an arbitrary solution to a problem, the algorithm then continuously makes small changes to the solution, selecting the neighboring one that provides the most improvement to the objective function. In our case we have utilised the the BIC Score.

Bayesian Information Criterion (BIC):

BIC is a statistical measure. It is utilised in order to evaluate the model fit of a set of data against the number of parameters in the model. The BIC balances the goodness of fit against the complexity (i.e. number of parameters) in the model.

$$BIC = -2 \times \ln(L) + k \times \ln(n) \quad (1)$$

Within the context of Bayesian Networks, lower BIC values tend to indicate better model structures. This is due to the fact that a lower BIC implies that the model describes the data well, but without being overly complex.

Maximum Likelihood Estimation (Parameter Estimation): Once out optimal structure has been determines, we can now work towards estimating the probabilities or parameters of the network. We have implemented this through the use of Maximum Likelihood Estimation (MLE).

$$P(X = x|pa(X) = u) = \frac{N(x, u)}{N(u)} \quad (2)$$

This method provides us with the probabilities that maximize the likelihood of observing given data. The model is essentially fit to the training data in the best way possible.

The implementation of a Bayesian Network requires us to find the balance between determining the right structure and estimating accurate probabilities. Our execution is a rigorous approach to ensure that our model is well-fitted, without risking the chance of under-fitting/over-fitting. Through the use of the pgmy libraries and the various approaches such as Hill Climbing, Bayesian Information Criterion and the Maximum Likelihood Estimation we were able to achieve a model which excels in both accuracy and complexity. We have successfully implemented a robust and generalized model.

The graph below represents our generated Bayesian Network. Each node represents a health metric/variable, and the edges capture the probabilistic relationships between these metrics.

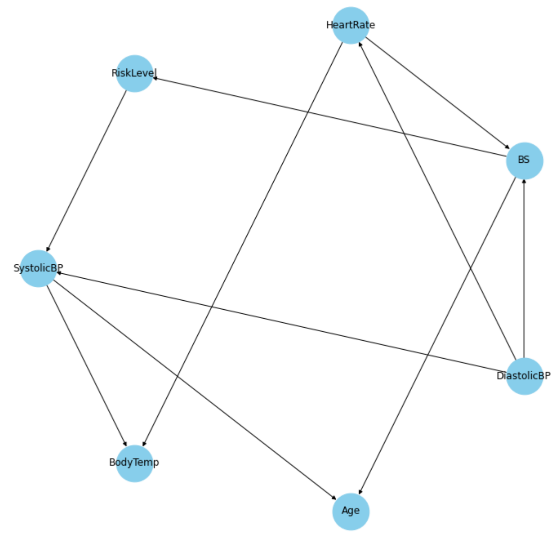

Fig. 2. Bayesian Network

C. Model Training, Hyper-parameter Tuning

Within this section we will discuss the methodology in terms of Model Training and Hyper-parameter Tuning.

1) Model Training: The structure of our model was determined using the methodology discussed in the previous section (Hill Climbing and Bayesian Information Criterion). The model is then trained on the training data which was split using StratifiedShuffleSplit as previously discussed. Once trained the model is then validated upon the validation data. The metrics that are then generated are based on the results of the models' outcome of the validation data.

2) Hyper-parameter Tuning: We have set a hyper-parameter,max_indegree. This represents the maximum number of parents a node can have in our Bayesian Network and limits the models complexity. The code evaluated various models with max_indegree values ranging between, and including, 1 to 5. This method is implemented and tuned to optimize our model's performance. The hyper-parameter is set to the value that provides the highest validation accuracy. The variance in validation accuracy over the various values of the hyper-parameter highlights the importance of hyper-parameter tuning in the model training process.

D. Evaluation

The precision of our model was measured through the use of 2 metrics: Area Under the Receiver Operating Characteristic Curve (AUC-ROC) and Logarithmic Loss (Log-Loss):

- AUC-ROC: This is a performance metric used to evaluate the capability of a binary classification model to discriminate between the positive and negative classes.

  ROC Curve: This graphical plot is utilised within the metric. The plot illustrated classification ability of the Binary Model as its discrimination threshold varies. The curve is plotted as True Positive Rate (TPR) against the False Positive Rate (FPR).

  AUC: This represents the are underneath the above described graphical plot. It then provides a single scalar value that represents the model's overall performance. A model with perfect classification would have a value of 1.

  Advantages: This metric is threshold-invariant thus it is able to evaluate out model's performance across all it's possible thresholds. In addition to this, it also takes into account both errors, false positives and false negatives.

- Log-Loss: Log Loss is a performance metric used to evaluate the prediction accuracy of a classification model. It is ideal in binary and multi-class classification models thus making it ideal for our model. The uncertainty of predictions is measured in terms of true labels. The Log-Loss binary classification formula is as follows:

$$-\frac{1}{N}\sum_{i=1}^{N}[y_i\log(p_i)+(1-y_i)\log(1-p_i)] \quad (3)$$

  Advantages: Incorrectness comes in varying types and degrees, this metric take this into account and has a penalizing scheme in place for both. A confidently wrong prediction penalized more than one that is merely uncertain.

  A perfect model would have a Log-Loss of 0.

AUC-ROC provides us with an an aggregate measure of a model's ability to distinguish between classes across thresholds, Log Loss, on the other hand, evaluates the accuracy of the predicted probabilities themselves. The combination of these metrics provide us insights into the confidence of predictions and a holistic view of our model performance.

1) Results: A confusion matrix was generated and is displayed below:

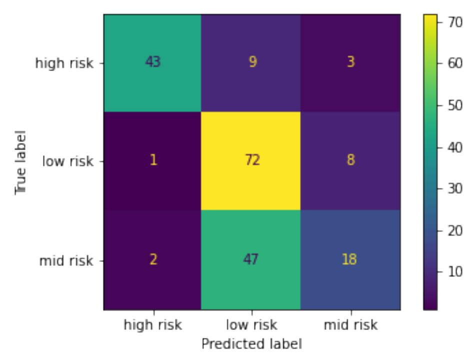

Fig. 3. Confusion Matrix

This confusion matrix was then used to calculate the models target values respective test accuracy, precision, recall, and F1 score.

TABLE II
PERFORMANCE METRICS

Risk Level	Precision	Recall	F1 Score
High Risk	0.93	0.78	0.85
Mid Risk	0.62	0.27	0.38
Low Risk	0.56	0.89	0.69

Our Bayesian Network was tasked with classifying entities into three risk levels: High Risk, Mid Risk, and Low Risk.

The models overall accuracy lies at 65.52%. This shows us that out of the total predictions made by the model on the testing/validation dataset 65.52% of these predictions were correct.

High Risk: In all instances that the model predicted as "High Risk," 93% of these predictions were correct. 78% of all "High Risk" instances were successfully identified.

The F1 score for this category lies relatively high, implying that this model provides good performance.

Mid Risk: In all instances that the model predicted as "Mid Risk," 62% of these predictions were correct. 27% of all "Mid Risk" instances were successfully identified. The F1 score for this category lies relatively low, implying that this model could use improvement in terms of training on this category.

Low Risk: In all instances that the model predicted as "Low Risk," 52% of these predictions were correct. 89% of all "Low Risk" instances were successfully identified. The F1 score for this category lies between the above two, this indicates a balnced performance

It is evident that out Bayesian network model performs best at classifying "High Risk" instances, in terms of both, high Precision and Recall. The model struggled with the "Mid Risk" category the most. This could be due to the fact that the category can be described as a bit of a gray area in terms of the data and is more difficult to medically diagnose and thus predict. Due to the great significance of the "High Risk" classification it is great that the model performs well in that aspect.

Further evaluation metric can be seen in the table below:

TABLE III
EVALUATION METRICS

Category	AUR-ROC	Log Loss
High Risk	0.94	0.25
Mid Risk	0.69	0.56
Low Risk	0.73	0.51
Overall	0.79	0.72

AUR-ROC :

High Risk: 0.94 indicates an excellent performance in discrimination capability between the positive class and the negative class within the High Risk category. *Mid Risk:* 0.69 suggests a fair performance, however it is notably lower than the prior category. *Low Risk:* 0.73 lies between the above two categories. This suggest that the model does what it is meant to with relatively good performance. *Overall:* The overall metric of 0.79, is a testament that showcases the models overall discrimination capability across all risk categories. It is important to note that the latter two categories do bring thus score down.

Log Loss:

High Risk: A log loss of 0.25 is low. This indicates the confidence of our model in terms of its predictions within the category. *Mid Risk:* The log loss of 0.56 is considerably higher than its prior. This indicates to us that the model deviates more from its true outcomes within said category. *Low Risk:* Consistent with the AUR-ROC metric, the log-loss value for this category lies between the 2 prior to it. This suggests a moderate level of confidence it its own predictions. *Overall:* The combined log loss of 0.72 further emphasizes that the model's performance is superior in the High Risk category as compared to the other risk groups.

Out Bayesian Model notably strong performance in predicting outcomes in the High Risk category. This fact remains consistent with all generated metrics. In contrast, the model's performance dips in the Mid and Low Risk categories. It is essential to consider the implications of the model's varying performance across risk categories in practical scenarios. Our model's superior performance in the High Risk category is reassuring, as this is the category that holds the most weight.

III. CONCLUSION

The healthcare field in general remains as an extremely volatile, taxing and resource intensive field. The maternal health sector plays an important role amongst us as humans as it can determine the outcome of future generations. Due to the implications it is evident that with involvement of the Computer Science field we can help advance the field.

It is apparent that our Bayesian Network performed well , specifically in classifying the 'High Risk' category. The inconsistencies, however noticeable should not be disheartening. Given the nature of the data there were various class imbalances and thus the model was more well trained on the 'High Risk' category. It is also important to note 'Mid Risk' category remains as a slight gray area medically. It is also important to note that the correct classification of the 'High Risk' category is significantly more important than that of the other two categories.

It would be interesting to see improvements added such as introducing Reinforcement Learning Techniques, data augmentation, and various other model training techniques. There is much room for advancement in the important field of Maternal Health.