AI in Medicine (SBE452)

4th SBME

Final project Report
(Fetal Health Classification)

By:

Hadeel Sameh Hassan

Samar Ibrahim Antar

Toaa Mahmoud Abdelhakeem

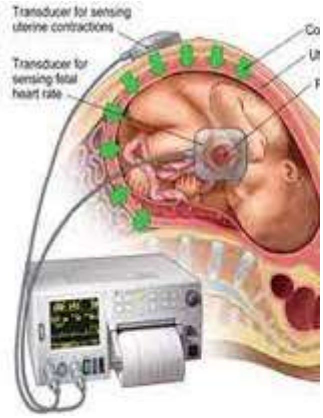
Ghada Adel Abdelaziz

Proposed to:

Dr. Enas Yassine

Eng.Christeen Ramsis

1)Problem description:



Cardiotocography (CTG) is a technical means of monitoring the baby's heart rate and the uterine contractions (UC) during pregnancy, are simultaneously recorded by means of two probes placed on the maternal abdomen, a US Doppler probe for fetal heart rate (FHR) signal and a pressure transducer for UC signal. Thus, CTG contains of two distinct signals: its continuous recording of instantaneous FHR and uterine activity (UC). The machine used to perform the monitoring is called a cardiotocograph, more commonly known as an electronic fetal monitor (EFM).The information which is acquired from CTG is used for early recognition of a pathological state of baby (i.e. congenital heart defect, fetal distress or hypoxia, etc.)and may help the obstetrician to predict future complications and interpose before there is a permanent damage to the fetus.

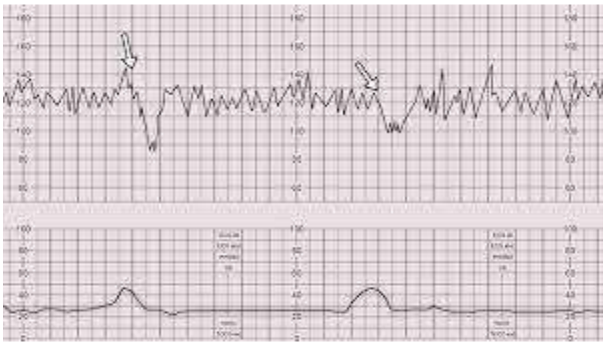
Therefore, we need to create multiclass classification models to classify CTG features into the three fetal health states :

1-Normal

2-Suspect

3-Pathological

2)Data and Features:



We have 21 features. Our features were extracted from CTG graph and features from 12 to 21 were extracted using the statistical way which use the signal in time domain and calculate Histogram then records it's Properties.

1. **Baseline Value**: The baseline value is approximation of the mean FHR rounded to increments of 5 beats per minute (bpm) during a 10-minute time window, excluding marked accelerations and decelerations and periods .
2. **Acceleration**: Accelerations are transient increases in FHR of 15bpm or more above the baseline and lasting 15 sec.
3. **Fetal movements.**
4. **Uterine Contractions.**
5. **Presence of Light decelerations:** It Results from the increased vagal tone due to compression of the fetal head during contractions.
6. **Presence of severe decelerations:** its type of variable deceleration and it is when the depth is below 70 bpm and the duration is longer than 60 seconds.Variable decelerations are often associated with a favorable outcome.
7. **Presence of prolonged decelerations:** prolonged decelerations Are defined as a decrease of FHR below the baseline of more than 15 bpm for longer than 90 seconds but less than 5 minutes.
8. **Number of Abnormal short term variability in FHR.**
9. **Mean value of short term variability**.
10. **Percentage time with abnormal long term variability in FHR.**
11. **Mean value of long term variability.**
12. **Histogram width.**
13. **Histogram min.**
14. **Histogram max.**
15. **Histogram number of peaks.**
16. **Histogram number of zeros.**
17. **Histogram mode:** The location of the highest peak of the histogram.
18. **Histogram mean.**
19. **Histogram median.**
20. **Histogram variance.**
21. **Histogram tendency:** where data in histogram presents the most.

3) The Models we employed :

We applied 4 Classification algorithms which are: KNN, ANN, SVM, and RandomForest. We also tested their results with and without normalization.

The Reasons we have chosen these models are:

KNN	SVM	ANN	RandomForest
KNN is effective for multiclass classification. That's first thing making us think to use KNN.	SVM take cares of outliers better than KNN.	These networks can learn from examples and apply them when a similar event arises, making them able to work through real-time events.	Random forest classifier will handle the missing values and maintain the accuracy of a large proportion of data.
Few hyper-parameters to tune (KNN mainly involves two hyper-parameters, K value & distance function.)	SVM is more effective in high dimensional spaces and is relatively memory efficient.	Even if a neuron is not responding or a piece of information is missing, the network can detect the fault and still produce the output.	If there are many trees, it won't allow over-fitting trees in the model.
	SVM works relatively well when there is a clear margin of separation between classes, so we did standardization, normalization and scaling on data to achieve that and did the same with the others models also to see how this affects.	They can perform multiple tasks in parallel without affecting the system performance, and this helps when having large data.	It has the ability to handle a large data set with higher dimensions with low cost.

4) Results:

- **KNN** was used with n_neighbors=10.
- **SVM** with decision_function_shape='ovo' built in method for making it multiclass classification.
- **ANN** with 3 hidden layers, they have the same number of nodes= 8, and solver='adam' for weight optimization with max_iterations=1000 & activation function= 'relu'.
- We applied standardization for **KNN,SVM,and ANN** by subtracting the mean from data, normalization and scaling gradually to see their influence together.
- **RandomForest** was employed with No.estimators(number of trees)=100 splitting criteria is using gini algorithm ,and max depth of a single tree =10.
- We applied **StandardScaler** module from **SikitLearn** to normalize data while employing **RnadomForest**.

The following table shows the accuracy for all of the employed models:

Models	Accuracy			
	Without standardization	With standardization	With normalization	With scaling between [1, -1]
SVM	0.84	0.91	0.925	0.925
ANN	0.798	0.873	0.887	0.915
KNN	0.887	0.887	0.906	0.906
RandomForest	0.9436	-	0.9451	-

We can decide based on our results that **RandomForest** classifier was the most effective algorithm for our problem.