6.Explainability of the Model

EXPLAINABLE AI(XAI)

Explainable AI refers to methods and techniques in the application of artificial intelligence technology (AI) such that the results of the solution can be understood by human experts. It contrasts with the concept of the "black box" in machine learning where even their designers cannot explain why the AI arrived at a specific decision. XAI is an implementation of the social right to explanation.

Using LIME-LIME stands for Local Interpretable Model Agnostic Explanation. The simplicity and use of LIME are its greatest strengths. Despite being extensive, LIME's main concept is quite clear-cut and straightforward. Let's explore what the name itself means first: Model agnosticism is a trait of LIME that allows it to provide justifications for any specific supervised learning model by considering it as a stand-alone "black box." LIME can therefore support practically any model that is currently in use. Local explanations refer to LIME providing justifications that are accurate in the immediate neighborhood of the observation or sample being explained.

Lime using ML Model	Decision Tree	Random Forest	Logistic Regression	Gaussian NB	SVC
Prediction Probabilities for Malignant	1.00	0.98	0.98	1.00	0.97
Bare Nuclei	0.41	0.32	0.32	0.28	0.33
Marginal Adhesion	0.13	0.26	0.27	0.24	0.26
Uniformity of Cell Shape	0.09	0.21	0.20	0.23	0.14
Mitoses	-0.06	-0.22	-0.22	-0.19	-0.31
Normal Nucleoli	-0.05	0.00	0.02	-0.05	0.02
Uniformity of cell Size	-0.04	-0.00	0.01	-0.04	-0.01
Single Epithelial cell Size	0.03	0.01	-0.01	-0.03	-0.00
Clump Thickness	0.03	0.04	0.03	-0.00	0.04
Bland Chromatin	0.00	0.01	0.02	-0.02	0.04

Table:7- Explaining the algorithms using LIME

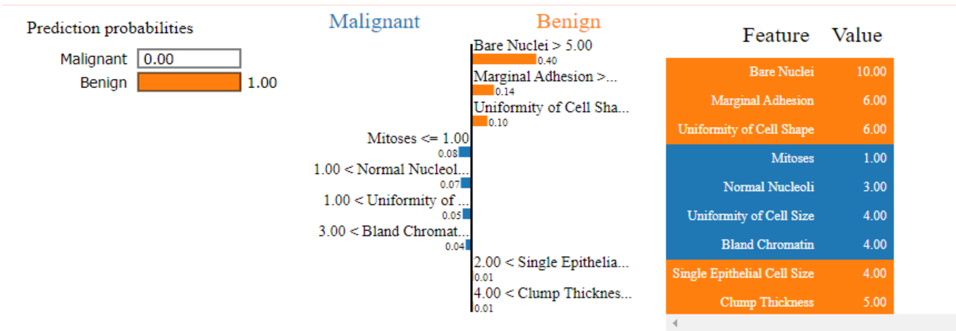

Fig-22 Explaining Decision Tree Using LIME

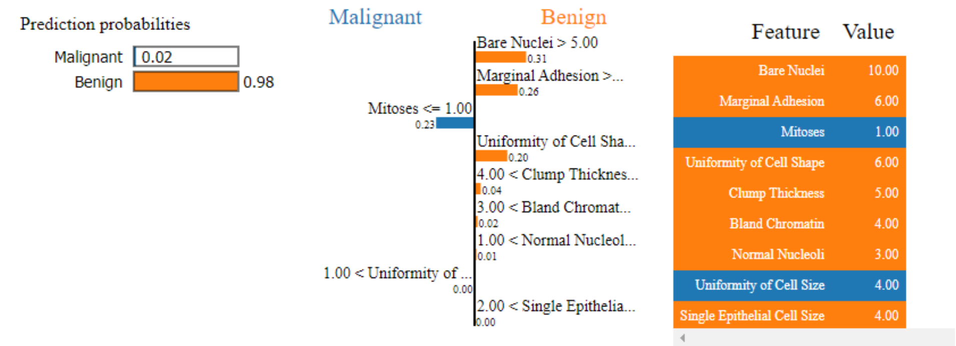

Fig-23 Explaining Random Forest Using LIME

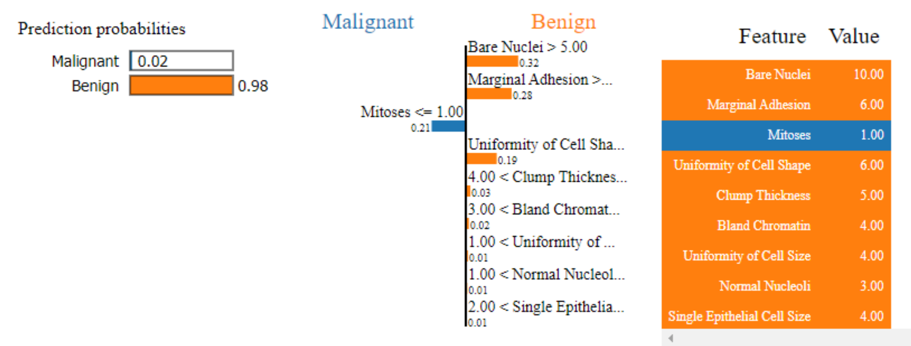

Fig -24 Logistic Regression Tree Using LIME

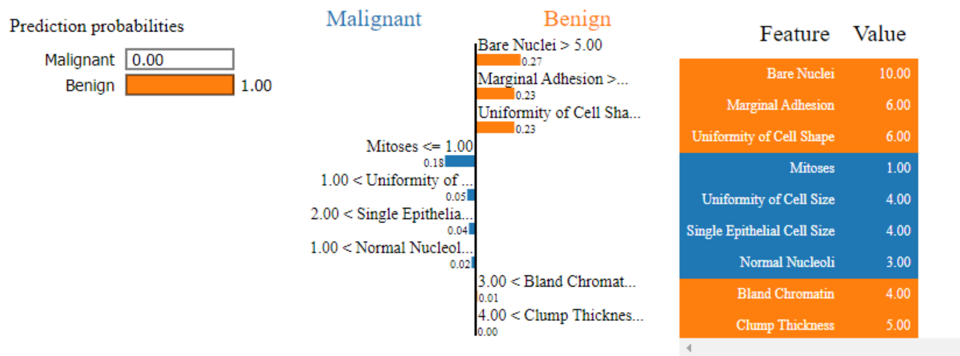

Fig -25 Gaussian Naive Bayes Tree Using LIME

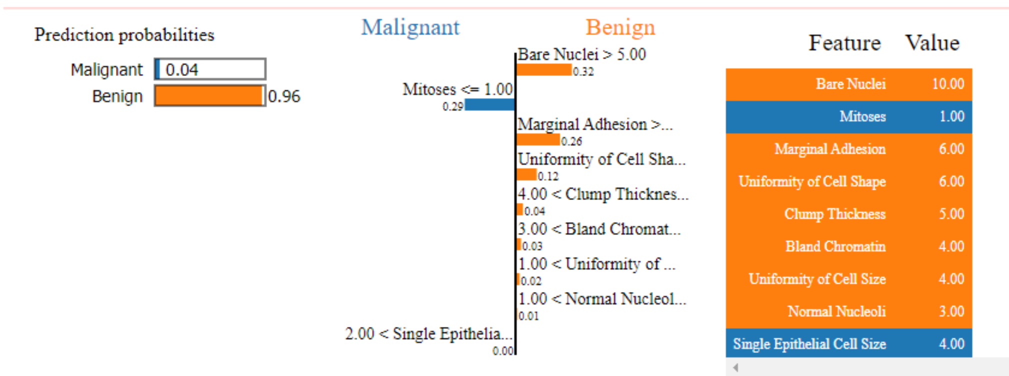

Fig-26 SVM Tree Using LIME

Using SHAP- SHAP is a mathematical method to explain the predictions of machine learning models. It is based on the concepts of game theory and can be used to explain the predictions of any machine learning model by calculating the contribution of each feature to the prediction.

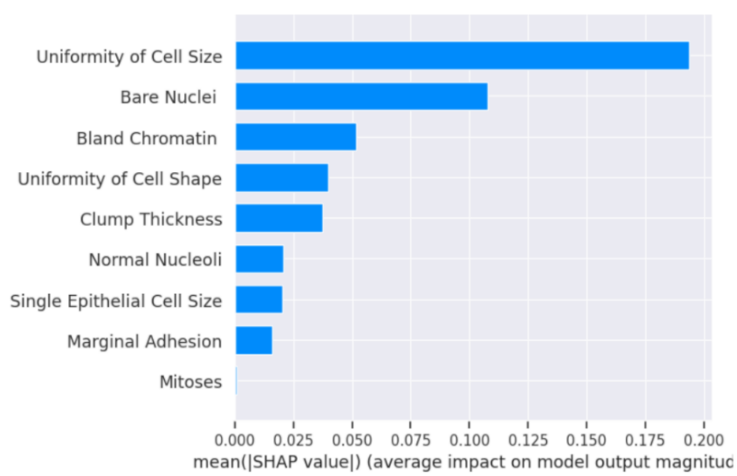

fig.27 Summary plot using SHAP.

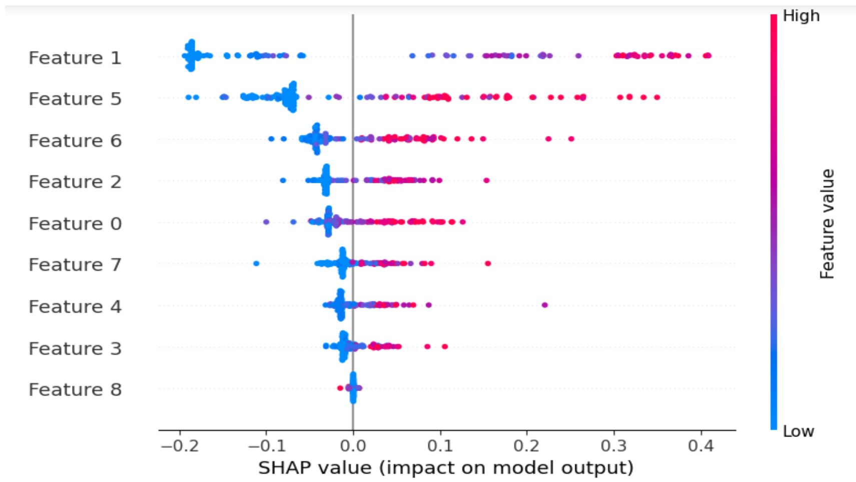

fig.28 Beeswarm plot using SHAP.

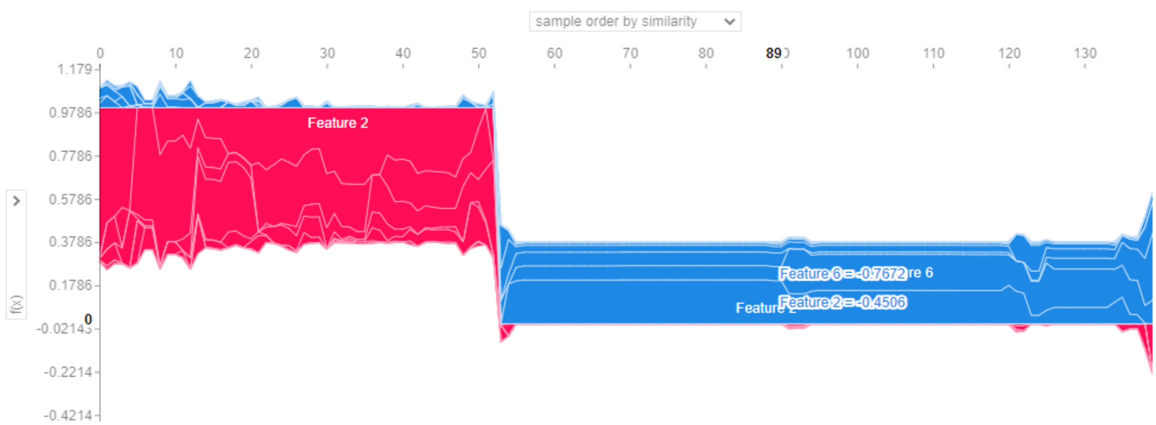

fig.29 force plot using SHAP for DecisionTreeClassifier

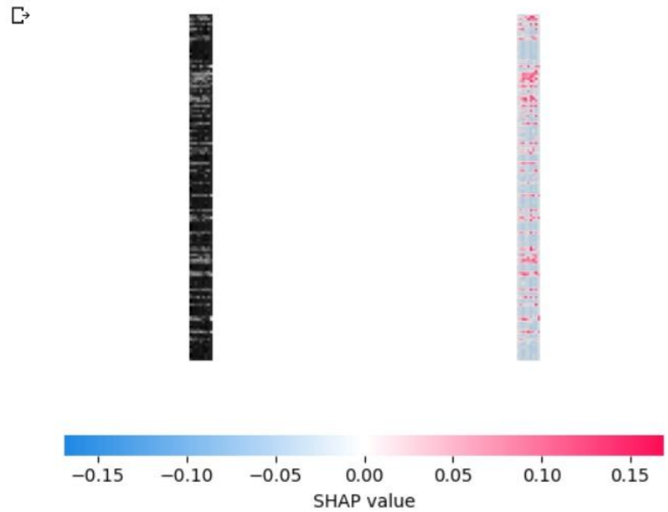

fig.30 Image plot of CNN

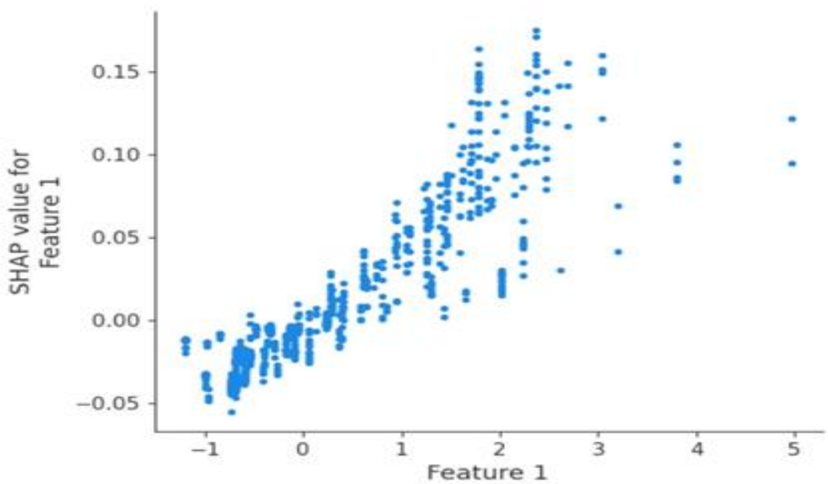

fig.31 Dependence plot of CNN

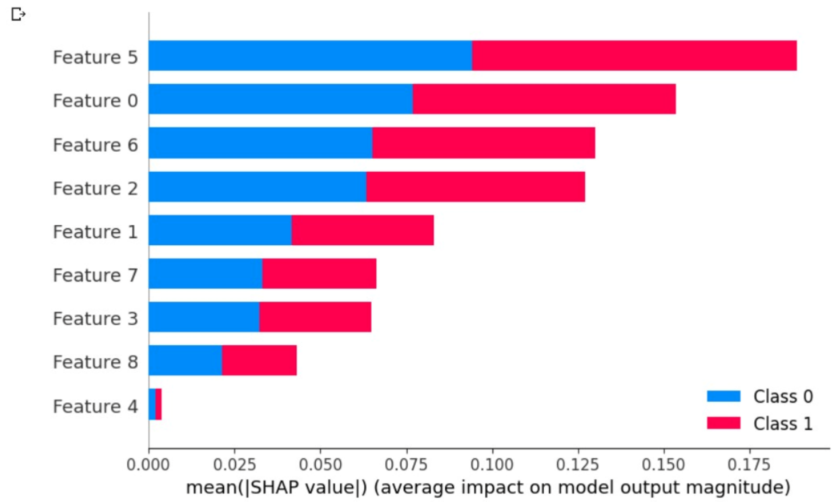

fig.32 Summary plot of FFNN

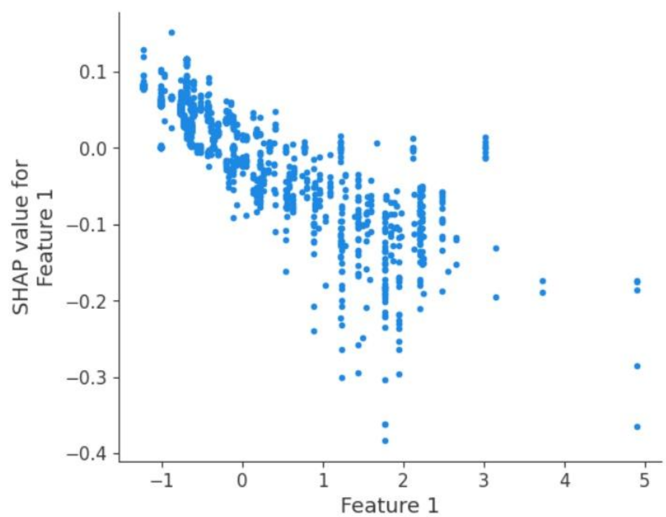

fig.33 Dependence plot of FFNN

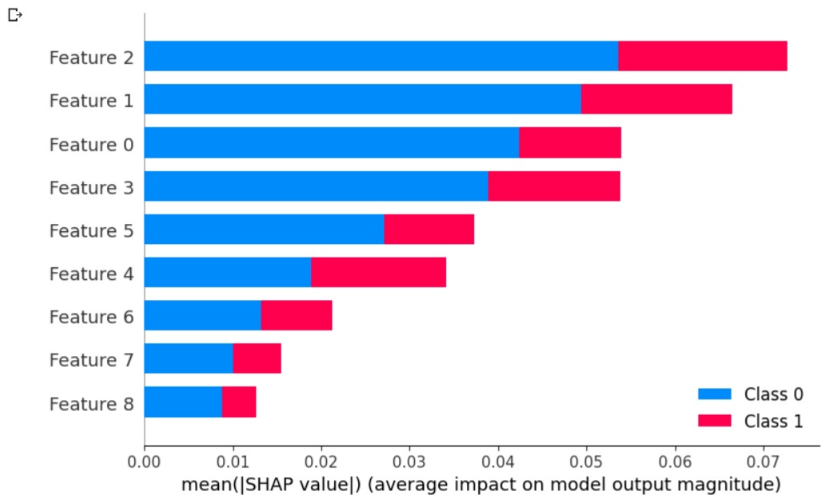

fig.35 Summary plot of RNN

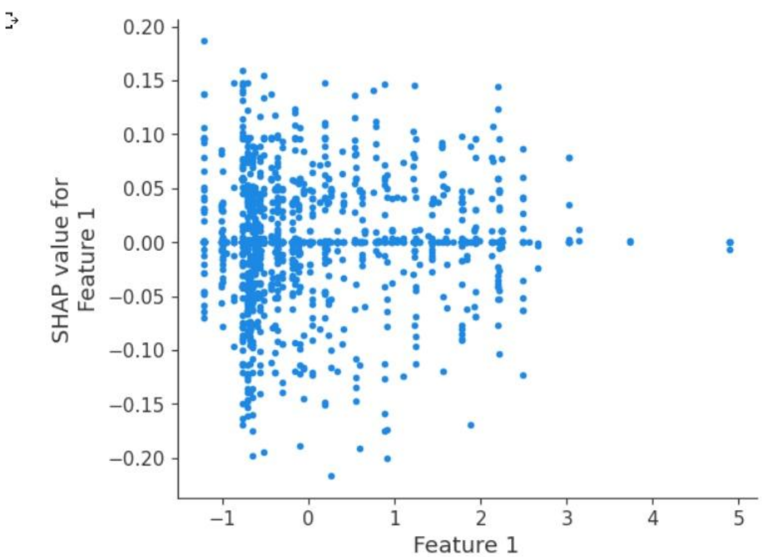

fig.36 Dependence plot of RNN