In searching for clinically reliable diagnosis methods to detect cancer in early stage, most effort had been focused on determining the ploidy status, which if properly handled, can serve as a reliable marker of cell proliferation. Recently, exfoliative cytology has emerged as a prominent technology in early oral cancer diagnosis. Although if offers a simple and non-invasive procedure, many technical hurdles largely limited this method from becoming an automated and robust clinical standard protocol. In this research, we explored a novel data analysis procedure and leveraged modern machine learning technique to efficiently utilize the DNA Index (D.I. value) obtained from the commercially available imaging analysis and successfully predicated disease outcomes. Using resampling methods for pruning the model core parameters, a final statistical model was determined. Our method showed high sensitivity and specificity application in both the training process and predicting on a hold-off test data. In the end, we proposed a risk index for the OLK patient, which has complex clinical aspects and provided a valuable guide for the clinical professionals for patient follow up schema.