

Classification of Cervical Cells Based on Extreme Gradient Boosting

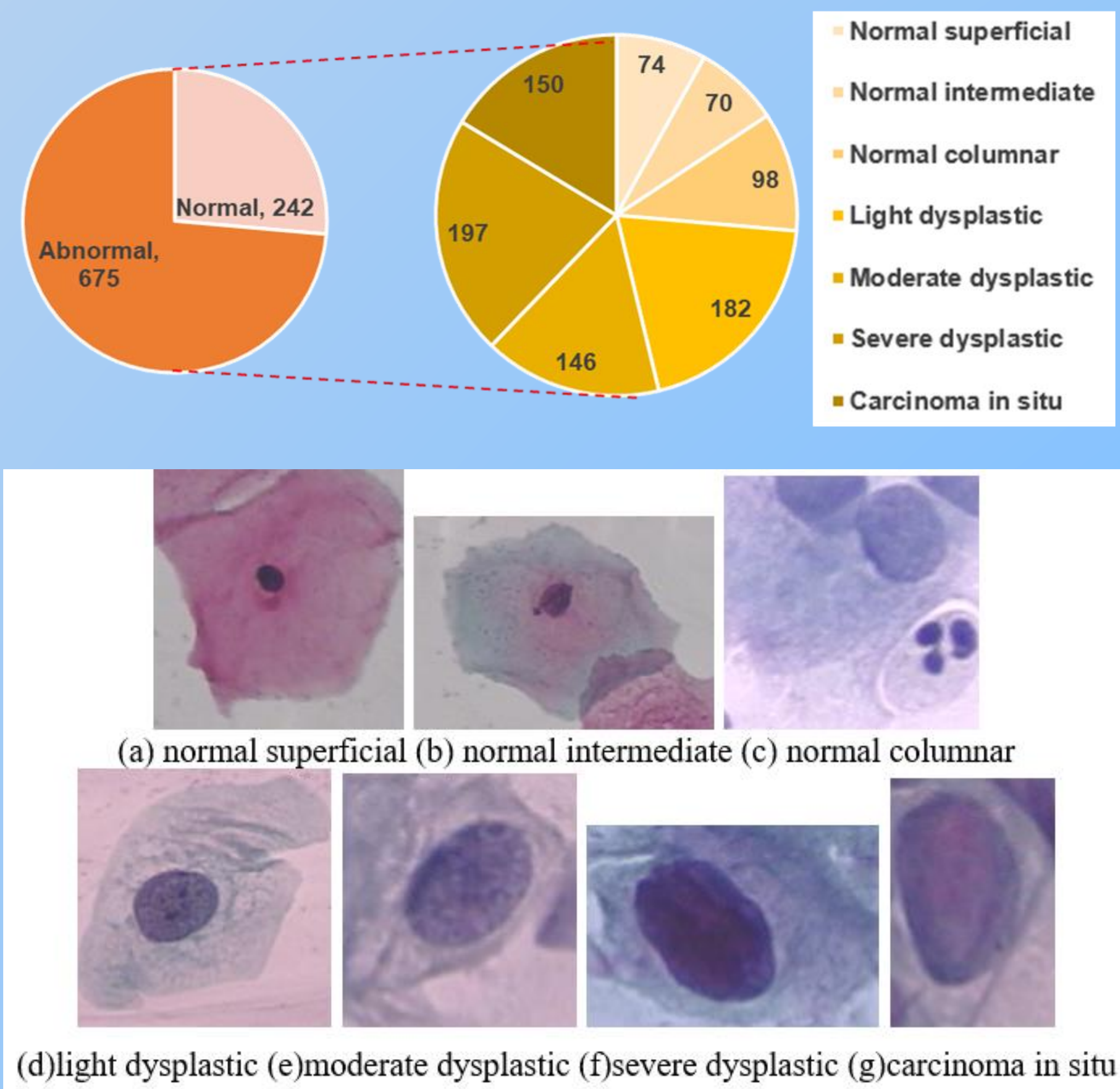
Lili Pei, Wei Li, Jiangang Ding, Ming Yang, Xingxing Sun
School of Information Engineering Chang'an University , Xi'an 710064, China
Email: peilili@chd.edu.cn ; grandy@chd.edu.cn

INTRODUCTION

Computer-aided diagnosis is invaluable in the treatment of cervical cancer. In view of the low accuracy of the existing methods for the refined automatic classification of cervical cell types, a cervical cell classification framework based on extreme gradient boosting (XGBoost) based on geometric feature is proposed. A data set of cervical cell characteristics is established by extracting cell morphological features and obtaining the principal components of the features through principal component analysis (PCA). The extracted features are classified using an extreme gradient boosting classifier based on cross-validation and grid search optimization. The accuracies of the two- and seven-category classifications are 0.9674 and 0.6594, respectively. This method provides a theoretical basis for the automatic detection of cervical cells.

Key words: cervical cell classification; image processing; deep learning; XGBoost.

METHOD

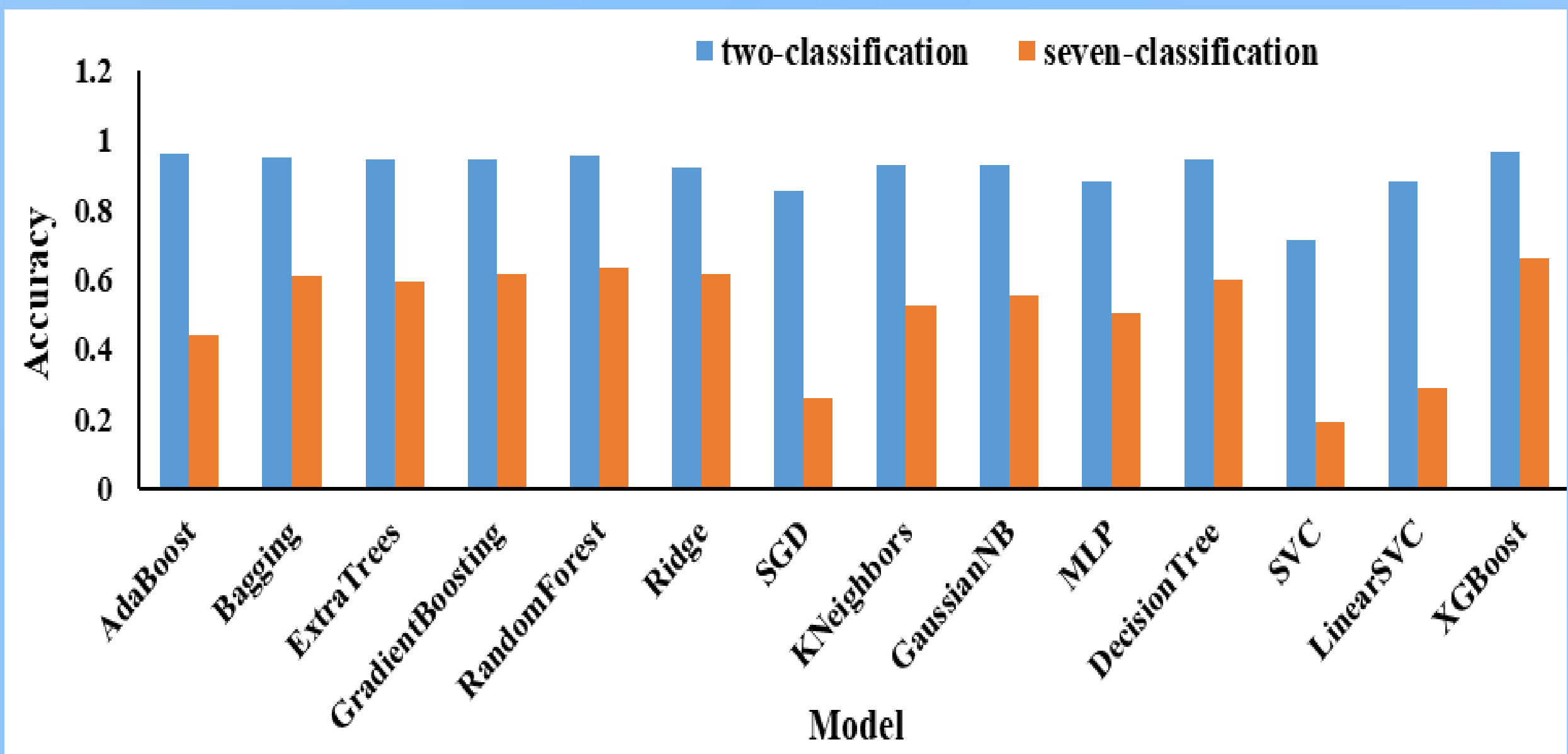


- A. Data preprocessing
- B. Cell morphological feature extraction
- C. Classification of cervical cells based on XGBoost

Taking the parameter values under the optimal combination, we used the XGBoost classifier to classify the cervical cells. The evaluation indicators of the two- and seven-classification model are shown in Table

Parameters	Default value	Search range	Final value 2- types	Final value 7- types
learning_rate	0.3	[0.01, 0.05, 0.07, 0.1, 0.2]	0.1	0.07
min_child_weight	1	[1, 6]	3	1
max_depth	6	[3, 10]	5	7
gamma	0	[0.1, 0.6]	0.5	0.3
subsample	0.6	[0.6, 0.9]	0.8	0.8
colsample_bytree	0.6	[0.6, 0.9]	0.8	0.8
reg_alpha	0.05	[0.05, 0.1, 1, 2]	0.05	0.1
reg_lambda	0.05	[0.05, 0.1, 1, 2]	1	1

RESULTS



- 1) The cervical cell image set is obtained by extracting the region of interest and image amplification. In addition, morphological feature extraction is performed based on digital image processing technology, and PCA dimensionality reduction is used to establish a cervical cancer cell feature dataset.
- 2) XGBoost was used to classify cervical cell images, and cross-validation and grid search methods were used to optimize the XGBoost model parameters. The XGBoost classifier outperformed other linear and integrated classifiers in cervical cell classification by achieving accuracies of 96.7% and 65.9% in the two- and seven-category data, respectively.