Ensemble Learning Techniques for Cervical Cytology Classification

Nguyen Thai Hoc

Institute of Applied and Technology,
University of Information and Communication Technology

July 16, 2024

Content

Introduction

Institute of Applied and Technology

Overview

- Cervical cancer originates in the lower part of the uterus and is often associated with viral infections -> carrying a high risk of transmission through sexual contact.
- It is the second leading cause of death from malignant diseases in women [2].
- The Pap smear test is a crucial screening procedure for identifying cancerous or precancerous cells in the cervix.
- In recent years, several research articles have explored the early detection of cervical cancer using machine learning techniques [1].
- Deep Learning have become popular in computer vision tasks -> demonstrating remarkable results compared to traditional machine learning algorithms -> performance comparable to that of clinical experts [5].

Overview (cont.)

=> This study analyzes the impact of ensemble learning techniques [3] on the classification performance of cervical cancer cell abnormality, comparing their effectiveness with individual models.
=> Explores how image size **128/224/256** influences model performance.

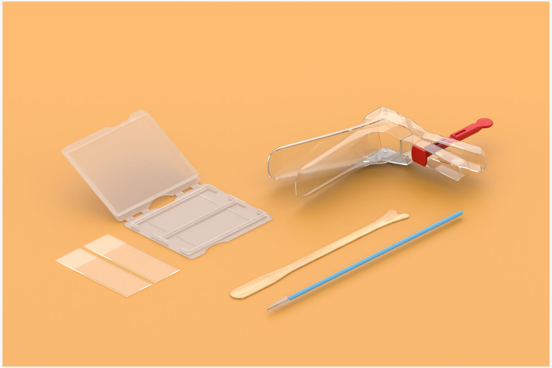

Figure: PAP smear

Datasets

Institute of Applied and Technology

Introduction of datasets

- Cervical cancer dataset from hospital A, Thai Nguyen, comprising **15,645** images of five common cervical cell abnormalities such as **ASC_US, LSIL, ASC_H, HSIL, and SCC**.

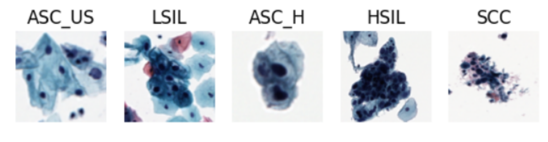

Figure: Five abnormal cell samples in the dataset.

- This dataset includes images of suspected diseased cells captured from various angles in variable-sizes.

Sampling and Preprocessing

- During preprocessing, utilized batch-wise image **augmentation** strategies.
- Specifically, within each training batch, images underwent a series of transformations, including **flipping**, **rotation**, and **adjustments in brightness**, **contrast**, and **scale**.
- Prior to model training, normalized all images to a pixel intensity range between 0 and 1 to standardize input data and facilitate effective learning.
- The dataset was splitted into three subsets: **training**, **validation**, and **test sets** with **80**%, **10**%, and **10**% of the data, respectivel.

Methodology

Institute of Applied and Technology

Classification Models

- The following architectures were chosen: **MobileNetV2** [10], **InceptionV2** [9], **InceptionV3** [8], **VGG16** [11], **ResNet101** and **Xception** [4]. These models were pretrained on the **ImageNet datasets** [7].
- The training process was conducted over **100 epochs**, use **Adam** optimizer configured with learning rate of **1e-4**.
- To prevent overfitting, **early stopping** and **regularization** were incorporated during the fine-tuning stages. The training was halted if no improvement in validation accuracy was observed for **10 consecutive epochs**.
- Set a **batch size of 16** -> striking a balance between computational efficiency and the stability of gradient descent.
- This approach aimed to leverage the strengths of pretrained models while fine-tuning them to effectively classify images of cervical cancer abnormality, ensuring robust performance across different architectures.

Ensemble Learning

- Ensemble Learning methods combine individual models, each generating separate predictions, to formulate a consolidated inference.

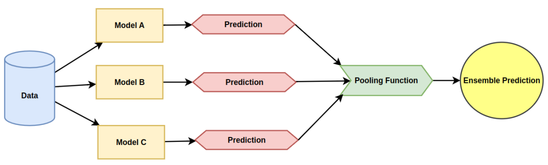

Figure: Ensemble Learning

- Stacking involves leveraging diverse base models and combining their outputs through methods such as voting or weighted averaging to produce a unified prediction.

Pooling Functions

- To aggregate ensemble predictions into a unified inference, investigated multiple methodologies and algorithms.
- Use **Voting**, **Fuzzy Distance** [6], **K-Nearest Neighbor**, **Naive Bayes**, **Decision Tree**, **Support Vector Machine**, and **Logistic Regression**.
- These methods aimed to integrate diverse predictions from multiple models into a unified and reliable prediction for our cervical cancer classification system.

Results

Single Models

Name	Accuracy	Precision	Recall	F1-Score
MobileNetV2	61.29	63.42	64.43	62.64
InceptionV3	**67.27**	**69.21**	**70.81**	**69.91**
InceptionResNetV2	64.22	66.60	67.83	66.80
VGG16	51.36	56.87	54.13	49.13
ResNet101	66.29	68.65	69.28	68.80
Xception	**70.00**	**71.90**	**72.19**	**71.91**

Table: Results single model for **input size 128**.

-> For an input size of 128, the **Xception** model exhibited the highest accuracy at **70.00%**, followed by **InceptionV3** and **ResNet101** with accuracies of **67.27%** and **66.29%**, respectively.

Single Models (cont.)

Name	Accuracy	Precision	Recall	F1-Score
MobileNetV2	65.51	68.84	69.05	68.41
InceptionV3	**69.93**	**73.24**	**72.25**	**72.57**
InceptionResNetV2	**71.62**	**74.36**	**74.12**	**74.04**
VGG16	61.94	65.22	65.15	64.65
ResNet101	58.83	61.08	61.91	60.41
Xception	65.58	67.74	68.75	67.18

Table: Results single model for **input size 224**.

–> For an input size of 224, **InceptionResNetV2** achieved the highest accuracy at **71.62%**, while **InceptionV3** closely followed with **69.93%**.

Single Models (cont.)

Name	Accuracy	Precision	Recall	F1-Score
MobileNetV2	65.58	67.83	68.81	67.62
InceptionV3	**70.12**	**72.37**	**72.57**	**72.30**
InceptionResNetV2	**68.76**	**72.10**	**70.65**	**71.12**
VGG16	67.20	69.18	70.41	69.75
ResNet101	63.57	66.58	67.58	66.43
Xception	67.27	68.66	70.92	69.39

Table: Results single model for **input size 256**.

For an input size of 256, **InceptionV3** showed the best performance with an accuracy of **70.12**%, while **InceptionResNetV2** and **Xception** achieved second and third highest accuracies of **68.76**% and **67.27**%, respectively.

Ensemble Models

Name	Accuracy	Precision	Recall	F1-Score
Voting	72.72	74.71	75.48	74.74
Fuzzy Distance	70.32	72.76	73.27	72.68
K-Nearest Neighbor	71.81	73.80	74.67	74.20
Naive Bayes	71.23	71.83	74.26	72.52
Decision Tree	70.97	73.42	73.99	73.67
Support Vector Machine	**73.11**	**74.86**	**75.83**	**75.29**
Logistic Regression	**73.63**	**75.49**	**76.38**	**75.89**

Table: Results ensemble model for **input size 128**.

-> **Logistic Regression** achieved the highest accuracy of **73.63%**, closely followed by **Support Vector Machine (SVM)** with an accuracy of **73.11%**.

Ensemble Models (cont.)

Name	Accuracy	Precision	Recall	F1-Score
Voting	**73.83**	**76.01**	**76.44**	**76.17**
Fuzzy Distance	72.59	75.31	75.28	75.20
K-Nearest Neighbor	73.05	75.32	75.43	75.36
Naive Bayes	73.44	74.19	76.29	75.01
Decision Tree	72.72	74.91	75.25	75.05
Support Vector Machine	**74.22**	**76.48**	**76.69**	**76.57**
Logistic Regression	73.76	76.05	76.23	76.14

Table: Results ensemble model for **input size 224**.

-> **SVM** achieved the highest accuracy of **74.22**%, followed by **Voting** and **Logistic Regression** with accuracies of **73.83**% and **73.76**%, respectively.

Ensemble Models (cont.)

Name	Accuracy	Precision	Recall	F1-Score
Voting	73.44	75.36	76.14	75.70
Fuzzy Distance	71.88	74.49	74.69	74.51
K-Nearest Neighbor	72.72	75.14	75.36	75.23
Naive Bayes	72.40	73.31	75.40	73.93
Decision Tree	70.06	72.64	72.91	72.76
Support Vector Machine	**73.83**	**76.31**	**76.33**	**76.29**
Logistic Regression	**73.83**	**76.26**	**76.34**	**76.29**

Table: Results ensemble model for **input size 256**.

-> **Voting**, **SVM**, and **Logistic Regression** all achieved the highest accuracy of **73.83%**.

Conclusion

Highlights

- Investigates ensemble learning techniques to improve the classification performance of medical images, focusing specifically on the challenge of identifying abnormalities in cervical cancer.
- The results highlight the effectiveness of ensemble learning in achieving notable performance gains and the influence of variable-size images on the model's performance.
- Uses stacking as the chosen ensemble learning method, comparing its performance against individual model training approaches.
- However, despite these advances, achieving optimal model accuracy remains a challenge, with some individual models exhibiting signs of overfitting, likely attributable to dataset quality and adequacy limitations.

References

References I

[1] Md Mamun Ali et al. "Machine learning-based statistical analysis for early stage detection of cervical cancer". In: *Computers in biology and medicine* 139 (2021), p. 104985.

[2] Marc Arbyn et al. "Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis". In: *The Lancet Global Health* 8.2 (2020), e191–e203.

[3] Mudasir A Ganaie et al. "Ensemble deep learning: A review". In: *Engineering Applications of AI* 115 (2022), p. 105151.

[4] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on CVPR*. 2016, pp. 770–778.

References II

[5] Kisuk Lee et al. “Superhuman accuracy on the SNEMI3D connectomics challenge”. In: *arXiv preprint arXiv:1706.00120* (2017).

[6] Rishav Pramanik et al. “A fuzzy distance-based ensemble of deep models for cervical cancer detection”. In: *Computer Methods and Programs in Biomedicine* 219 (2022), p. 106776.

[7] Olga Russakovsky et al. “Imagenet large scale visual recognition challenge”. In: *International journal of computer vision* 115 (2015), pp. 211–252.

[8] Mark Sandler et al. “Mobilenetv2: Inverted residuals and linear bottlenecks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 4510–4520.

References III

[9] Christian Szegedy et al. "Rethinking the Inception Architecture for Computer Vision". In: ().

[10] Mingxing Tan and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks". In: *ICML*. PMLR. 2019, pp. 6105–6114.

[11] Saining Xie et al. "Aggregated residual transformations for deep neural networks". In: *CVPR*. 2017, pp. 1492–1500.

Future Works

The next 3-6 months future works

- Continue to conduct in-depth research on combination learning techniques for cervical cancer classification problems to improve performance and accuracy, and to enable practical implementation.
- Develop multiple problems and approaches around the dataset to increase diversity.
- Research solutions for data preprocessing and handling to improve model performance.

Proposal

Institute of Applied and Technology

Proposal

- Invest in additional equipment such as Colab accounts and large Kaggle datasets to test larger models and handle more data.
- Reconfigure the institute's workspace in a European style to create a conducive atmosphere and inspire research and work.
- Enhance offline activities in two areas: First, by holding more seminars, and second ”ub” or ”ur” to exchange ideas, discuss, and strengthen relationships between the institute's leadership and colleagues.

Thank you everyone for your concern !

Nguyen Thai Hoc

Institute of Applied and Technology,
University of Information and Communication
Technology

July 16, 2024