BRITISH JOURNAL OF HOSPITAL MEDICINE

ARTICLE

Utilising deep learning networks to classify ZEB2 expression images in cervical cancer

Chenyang Zheng[1]

Qinqin Shen[1]

Lingjun Zhao[1]

Yijun Wang[1]

Author details can be found at the end of this article

Correspondence to:
Yijun Wang
(wangyijun5588@sina.com)

Abstract

Aims/Background Cervical cancer continues to be a significant cause of cancer-related deaths among women, especially in low-resource settings where screening and follow-up care are lacking. The transcription factor zinc finger E-box-binding homeobox 2 (ZEB2) has been identified as a potential marker for tumour aggressiveness and cancer progression in cervical cancer tissues.

Methods This study presents a hybrid deep learning system developed to classify cervical cancer images based on ZEB2 expression. The system integrates multiple convolutional neural network models—EfficientNet, DenseNet, and InceptionNet—using ensemble voting. We utilised the gradient-weighted class activation mapping (Grad-CAM) visualisation technique to improve the interpretability of the decisions made by the convolutional neural networks. The dataset consisted of 649 annotated images, which were divided into training, validation, and testing sets.

Results The hybrid model exhibited a high classification accuracy of 94.4% on the test set. The Grad-CAM visualisations offered insights into the model's decision-making process, emphasising the image regions crucial for classifying ZEB2 expression levels.

Conclusion The proposed hybrid deep learning model presents an effective and interpretable method for the classification of cervical cancer based on ZEB2 expression. This approach holds the potential to substantially aid in early diagnosis, thereby potentially enhancing patient outcomes and mitigating healthcare costs. Future endeavours will concentrate on enhancing the model's accuracy and investigating its applicability to other cancer types.

Key words: Cervical cancer; Convolutional neural network (CNN); Grad-CAM visualization; ZEB2 expression

Submitted: 07 April 2024; Revised: 15 May 2024; Accepted: 20 May 2024

Introduction

Cervical cancer (CC) is a malignant tumour originating from cervical cells, primarily associated with Human Papillomavirus (HPV) infection and characterised by early-stage precancerous lesions (Bray et al, 2018; Cohen et al, 2019). It disproportionately affects women in underdeveloped and developing countries, where 85% of cases are detected due to limited medical resources and inadequate screening programmes (Small et al, 2017; Hull et al, 2020). Routine screenings, crucial for early detection, hinge on the availability of facilities, test accuracy, and prompt follow-up (Cox et al, 2012; Kessler, 2017). The ThinPrep Cytology Test (TCT), offering superior performance compared to the conventional Pap smear, facilitates efficient and precise cell collection from the cervix and can detect HPV infections (Gibb and Martens, 2011). Widespread HPV vaccination and comprehensive screening initiatives could potentially reduce CC incidence by up to 90% (Eddy, 1986; Demarco et al, 2020; Arezzo et al, 2021).

Diagnosis typically involves Pap or HPV tests and, if suspicious, a biopsy (Jusman et al, 2014). Manual analysis of traditional cytology is prone to errors, but automatic image recognition technology now offers more consistent and reliable results (Zhang and Liu, 2004). Recent research suggests that high zinc finger E-box-binding homeobox 2 (ZEB2) expression in cancer tissues may promote tumour aggressiveness and metastasis (Singh et al, 2008; Li et al, 2017). ZEB2, a pivotal regulator in cellular processes such as epithelial-to-mesenchymal transition, potentially serves as a non-invasive biomarker for

How to cite this article:
Zheng C, Shen Q, Zhao L, Wang Y. Utilising deep learning networks to classify ZEB2 expression images in cervical cancer. Br J Hosp Med. 2024. https://doi.org/10.12968/hmed.2024.0156

cancer diagnostics and progression monitoring, given its role in enhancing migration and invasion while reducing cell adhesion (Diao et al, 2019; Feng et al, 2019; Ye et al, 2019).

In recent years, various forms of artificial intelligence, including Machine Learning (ML), Deep Learning (DL), and Machine Vision (MV), have been employed in the analysis of cervical histopathological images. The objective is to assist experts in obtaining quicker, more reliable, unbiased, and quantitative results (Tseng et al, 2014; Li et al, 2020; Chandran et al, 2021; Xue et al, 2022). In 2017, Ceylan et al proposed a multi-label classification technique for diagnosing early CC (Ceylan and Pekel, 2017). The same year, Taha et al (2017)introduced an image recognition system to detect CC in Pap smears. Furthermore, two popular voting classifiers and DNN classifiers in ML have been utilised to predict CC (Rayavarapu and Krishna, 2018). In 2018, William and colleagues presented an innovative use of Pap smear images to automatically identify CC (William et al, 2018). This approach was developed considering that changes in the genetic makeup caused by viruses can lead to CC, as noted in a study by Tian et al (2019). To enable this method, the researchers created a specialised capture panel covering 39 Mb, focusing on 17 types of HPV and 522 genes with known mutations linked to CC. With the continuous advancement of ML algorithms, an increasing number of researchers are adopting automated diagnosis systems for CC.

A novel method for identifying and categorising CC in Pap smear images was recently introduced by Khamparia et al (2020) using an Internet of Health Things-based DL framework that employs transfer learning techniques (Khamparia et al, 2020). In parallel, Ijaz et al (2020) proposed a CC prediction model that incorporates risk factors to predict the disease at an early stage. Researchers Toratani et al (2018) and Kang et al (2023) evaluated the accuracy of image recognition using convolutional neural networks (CNNs) in identifying microscopic images, yielding promising outcomes (Toratani et al, 2018; Kang et al, 2023). Despite their significant potential, these models face notable limitations that impede their practical application. Specifically, they often encounter challenges related to data imbalance, which can affect their predictive accuracy. Additionally, there is a risk of overfitting, especially in models trained on smaller, less diverse datasets. Furthermore, improvements are needed in their generalisation capabilities to ensure consistent performance across various clinical environments.

In recent years, ML and DL techniques have proven effective in identifying CC, primarily focusing on classifying CC electronic medical record data or Pap smear images. This study seeks to advance the field by introducing a hybrid system based on DL Networks such as EfficientNet, DenseNet, and InceptionNet to classify CC medical images based on the expression of the transcription factor ZEB2, which is associated with cancer progression. Additionally, the study uses GradCAM technology to visualise the regions of an image that are most crucial in classification decisions, offering insights into the features utilised by a DL system for classification. The primary motivation for this research is to provide more accurate and efficient methods for diagnosing CC, ultimately enhancing patient outcomes and reducing healthcare costs.

Methods

Proposed deep learning system

The CNN-based hybrid model proposed in this study for classifying ZEB2 expression images in CC follows the flowchart depicted in Figure 1. Initially, ZEB2 expression images are obtained, and preprocessing techniques are applied to enhance their quality. Subsequently, the preprocessed inputs are fed into three distinct CNN models—EfficientNet, DenseNet, and InceptionNet—selected for their superior performance in image classification tasks. Each CNN model independently processes the images and generates its predictions. These predictions from all three models are then combined using an ensemble voter, which considers the outputs of each model to make a final decision based on their collective class performance. The gradient-weighted class activation mapping (Grad-CAM) method is employed to highlight the most crucial parts of the image that influenced the classification outcome, providing valuable insights into the characteristics utilised by the hybrid model to identify ZEB2 expression in CC images. Overall, the proposed CNN-based hybrid model

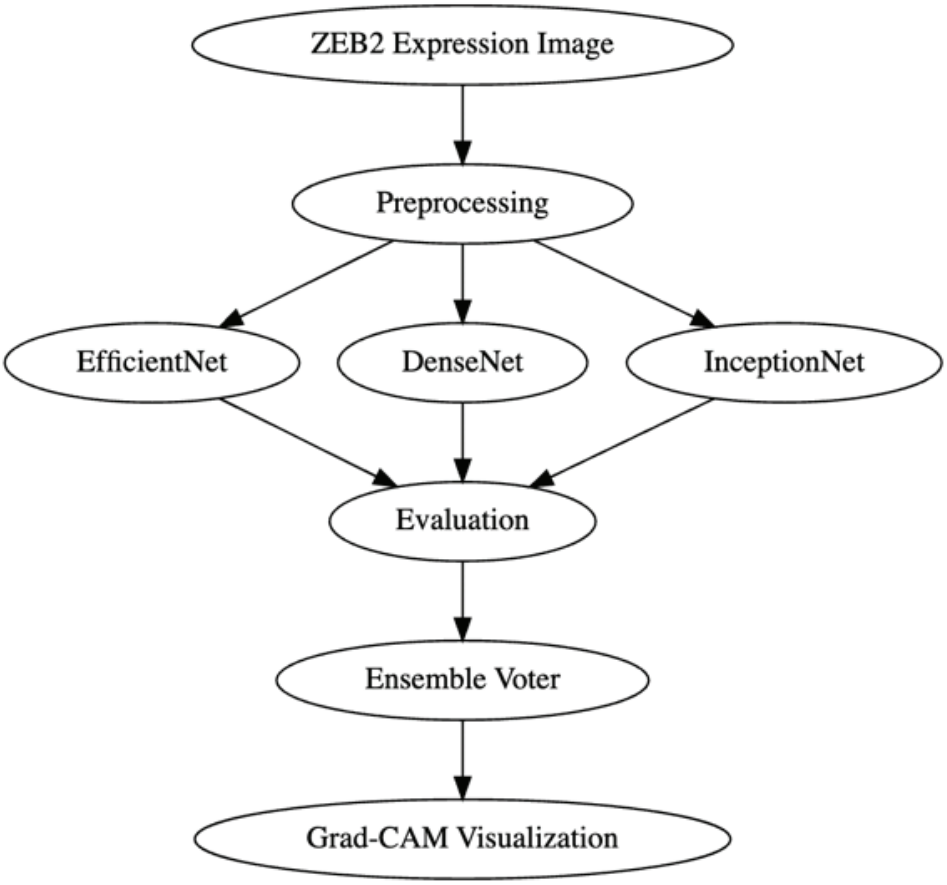

Figure 1. Flowchart of a convolutional neural network-based hybrid model for classifying zinc finger E-box-binding homeobox 2 (ZEB2) expression images in cervical cancer. Grad-CAM, gradient-weighted class activation mapping.

offers a comprehensive and effective solution for classifying ZEB2 expression images in CC. The integration of multiple CNN models and the ensemble voting strategy enhances the accuracy and robustness of the classification, while the Grad-CAM visualisation ensures interpretability and transparency in the decision-making process.

Data collection and pre-processing

The dataset utilised in this study was sourced from a hospital and comprises 649 images depicting ZEB2 expression in CC. We confirm that all procedures adhered to ethical standards established by the responsible committee on human experimentation (institutional and national), as well as the Helsinki Declaration of 1964, revised in 2013. Explicit consent was obtained from all human subjects included in this study for the publication of their images. Before obtaining consent, participants were thoroughly informed about the objectives, scope of the study, potential risks, and benefits involved. Following analysis, it was determined that the positive and negative samples in the ZEB expression image database totalled 260 and 389, respectively, resulting in a ratio of 0.67:1. The positive and negative distributions within the two databases are unbalanced, which could potentially impact the training accuracy of the model. Hence, data augmentation is deemed necessary.

Each image has a resolution of 986 × 1920 pixels and encompasses various types of cervical cells, alongside other cells such as white blood cells. Medical practitioners categorised all visuals into three classes: normal, low-grade, and high-grade. However, for the purpose of this research, these groups were consolidated into normal and abnormal, with the latter encompassing both low-grade and high-grade categories. The dataset was divided into three subsets for training (60%), validation (20%), and testing (20%).

In this study, several preprocessing techniques were employed to prepare the ZEB2 expression images in CC for classification using CNN. These techniques encompassed

image resizing, adding greyscale masks, incorporating Gaussian noise, and employing data augmentation. Image resizing was conducted to ensure uniform dimensions across all images, facilitating their proper input into CNN models. The images were resized to a standard resolution of 360 × 360 pixels, a common practice in image classification tasks. To further refine image quality and eliminate irrelevant features that could potentially interfere with the classification process, greyscale masks were applied. These masks were generated by thresholding the original images, retaining only regions with high-intensity values corresponding to cell nuclei. This approach directed the network's focus towards the most critical areas of the images, likely containing pertinent features for ZEB2 expression classification. Additionally, Gaussian noise was introduced to the images to mimic real-world noise and enhance the robustness of the CNN models. A minor amount of stochastic Gaussian noise was infused into the pixel values of the images, aiding in generalisation and preventing overfitting. To diversify the training data and mitigate overfitting, data augmentation techniques (Shorten and Khoshgoftaar, 2019) were implemented. These techniques included random rotations, horizontal and vertical flips, and random zooms. Random rotations simulated various angles of image acquisition, enhancing the model's adaptability to different orientations. Figure 2 illustrates the impact of image resizing, greyscale mask application, and Gaussian noise addition on image comparison. Figure 3 showcases data augmentation operations, including contrast adjustments, cropping, and rotation.

Classification with the convolutional neural network

In this study, three convolutional neural network (CNN) models were employed to classify ZEB2 expression images in CC. These models included DenseNet169, Inception V4, and EfficientNet-B7. Previous research has demonstrated the success of these models in tasks such as image classification and recognition, leveraging distinct architectural features to achieve high accuracy with reduced parameter count (Zhong et al, 2020; Khan et al,

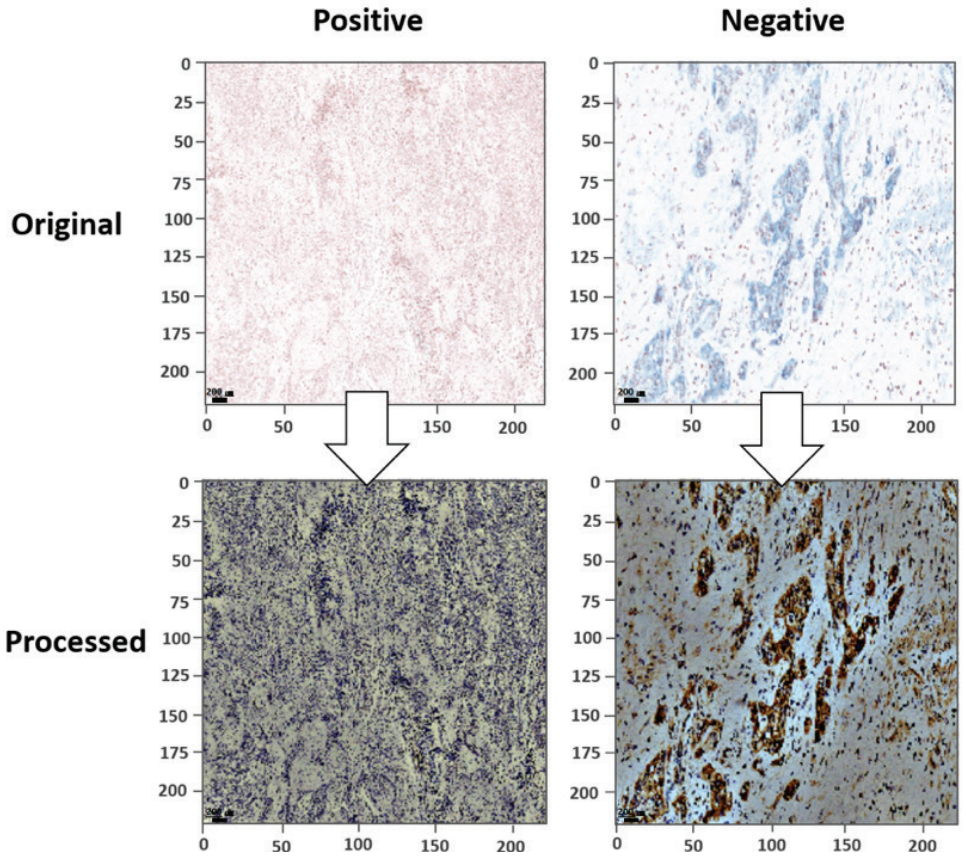

Figure 2. ZEB2 expression images comparison before and after resizing the image, greyscale mask application, and Gaussian noise addition. Scale bar=200 μm.

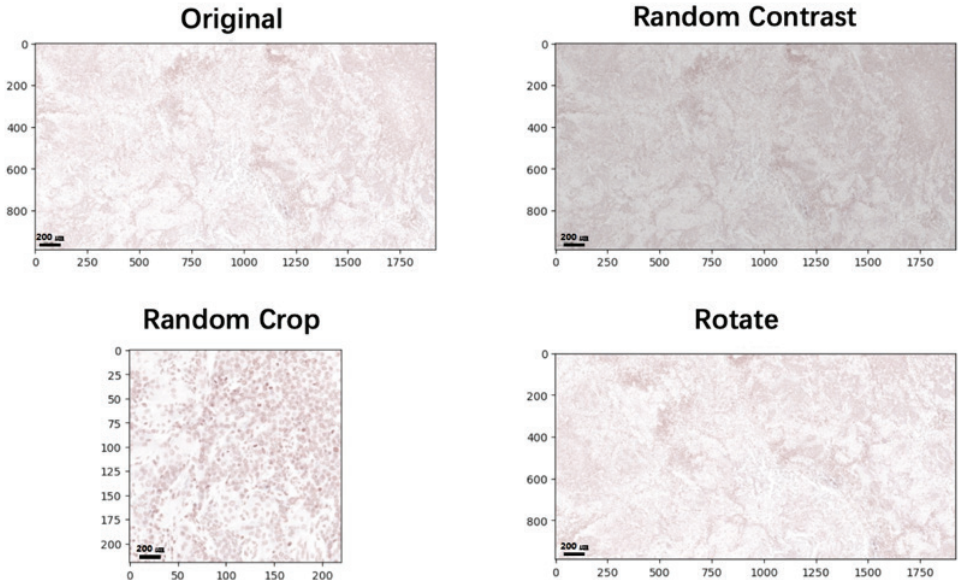

Figure 3. Data augmentation of ZEB2 expression images including contrast adjustments, cropping, and rotation. The units on both horizontal and vertical axes are in pixels. Scale bar=200 µm.

2021; Al Husaini et al, 2022). DenseNet169, for instance, is characterised by densely connected blocks of layers, minimising parameters while maintaining accuracy (Iandola et al, 2014). Its architecture consists of dense blocks, each integrating convolutional, batch normalisation, and ReLU layers, facilitating the reuse of features across layers and parameter reduction. Transition blocks following dense blocks further decrease spatial dimensions via convolutional, batch normalisation, and pooling layers. This approach enables each layer to use the feature maps of all previous layers, which promotes the reuse of features and reduces the number of parameters. Additionally, DenseNet169 has a transition block following each dense block, which includes a convolutional layer, batch normalisation, and a pooling layer to decrease the spatial dimensions of the feature maps. Similarly, Inception V4 is another CNN architecture that was created to attain high accuracy while using fewer parameters. The architecture consists of multiple inception modules, where each module is composed of parallel convolutional branches with different kernel sizes and pooling layers (Szegedy et al, 2017). The outputs of each branch are concatenated and fed into the next module. Inception V4 also incorporates residual connections between modules to facilitate the flow of gradients and prevent the vanishing gradient problem. The architecture also includes a stem module at the beginning, which consists of multiple convolutional and pooling layers to extract low-level features from the input images. EfficientNet-B7 was developed using neural architecture search to optimise the model's accuracy and efficiency. The architecture consists of multiple blocks, where each block is composed of a series of convolutional, batch normalisation, and activation layers, followed by a squeeze-and-excitation layer to emphasise important features (Tan and Le, 2019). Mobile inverted bottleneck blocks are utilised to connect the blocks, decreasing the feature map's spatial dimensions. EfficientNet-B7 adopts a compound scaling technique to enhance the model's depth, width, and resolution, which results in achieving greater accuracy with fewer parameters.

During the training phase, the preprocessed images were inputted into each model. Model parameters were then optimised to minimise classification loss utilising the Adam optimiser with a learning rate set to 0.001. The training was conducted for 20 epochs, and based on validation loss metrics, the weights of the best-performing model were saved.

Performance evaluation

The study assessed the performance of three CNN models in classifying ZEB2 expression in CC images, utilising various technical indicators. Accuracy was evaluated by dividing the

number of correctly classified images by the total number of test set images. Additionally, the F1 score was calculated, which considers both precision and recall. Precision represents the true positive rate divided by the sum of true and false positives, while recall is the true positive rate divided by the sum of true positives and false negatives. The F1 score ranges from 0 to 1, with higher scores indicating better performance. In addition to accuracy and F1 score, the receiver operating characteristic (ROC) curve and area under the curve (AUC) were employed to assess the CNN models. The ROC curve illustrates the trade-off between true positive rate (TPR) and false positive rate (FPR) across various threshold settings. TPR is the ratio of true positives to all positives, while FPR is the ratio of false positives to all negatives. The AUC, a scalar metric, measures the area under the ROC curve. A perfect classifier achieves an AUC of 1, while a random classifier achieves an AUC of 0.5. Each model underwent testing using a distinct validation dataset, with data partitioned into training and validation sets at a ratio of 80:20. Models were trained on the training set and evaluated on the validation set, with this process repeated for each model. Results were then compared to determine the best-performing model.

Ensemble voter

After training the three CNN models, DenseNet169, Inception V4, and EfficientNet-B7, the next step was to integrate them into a single model using an ensemble voter. The ensemble voter is an ML technique provided by the 'sklearn' library that takes the predictions from multiple models and combines them into a single final decision. The voter considers the class probabilities predicted by each model and aggregates them into a final class prediction. The ensemble voter can be configured to use a variety of methods to combine the predictions, including simple averaging, weighted averaging, or choosing the most frequent prediction. In this study, we used a simple averaging method to combine the predictions from the three models. The ultimate forecast was produced by taking the average of the class probabilities that were projected by every model. Earlier research indicated that this ensemble voting approach can enhance the overall precision of the model and decrease the variability in the predictions (Kannojia and Jaiswal, 2018; Ko et al, 2019).

Grad-CAM visualisation

To comprehend the characteristics that the three CNN models were acquiring while being trained, we utilised a technology called Grad-CAM (Gradient-weighted Class Activation Mapping) (Selvaraju et al, 2017) to display the activation maps of each model's last convolutional layer. Grad-CAM is a visualisation technique that emphasises the portions of an image that contribute significantly to a DL model's specific class prediction. It creates a heatmap signifying the importance of each pixel in the input image for a particular class prediction. The method calculates the gradient of the forecasted class score in relation to the feature maps of the last convolutional layer to generate these heatmaps. These gradients are subsequently employed to determine the significance of each feature map for the given class prediction. Finally, the importance weights are utilised to create a weighted combination of the feature maps, which generates the heatmap.

Results

The accuracy convergence of the training process for the three CNN models to classify ZEB2 expression images in CC is shown in Figure 4. The accuracy of the models plateaued around the fifteenth epoch, indicating that the models had converged. However, it is worth noting that the models may suffer from overfitting if the training continues beyond the point of convergence. EfficientNet-B7 had the highest accuracy throughout the training process, with an accuracy of 0.68 in the first epoch and reaching an accuracy of 0.97 after the twentieth epoch. Inception V4 started with an accuracy of 0.55 in the first epoch and reached an accuracy of 0.94 after the twentieth epoch. DenseNet169 started with an accuracy of 0.6 in the first epoch and reached an accuracy of 0.92 after the twentieth epoch. Table 1 presents the evaluation indicators for the three CNN models—DenseNet169, Inception V4, EfficientNet-B7, and the Ensemble classifier that were used to classify ZEB2 expression images in CC. The evaluation metrics include accuracy, AUC, and F1 score. Based on the

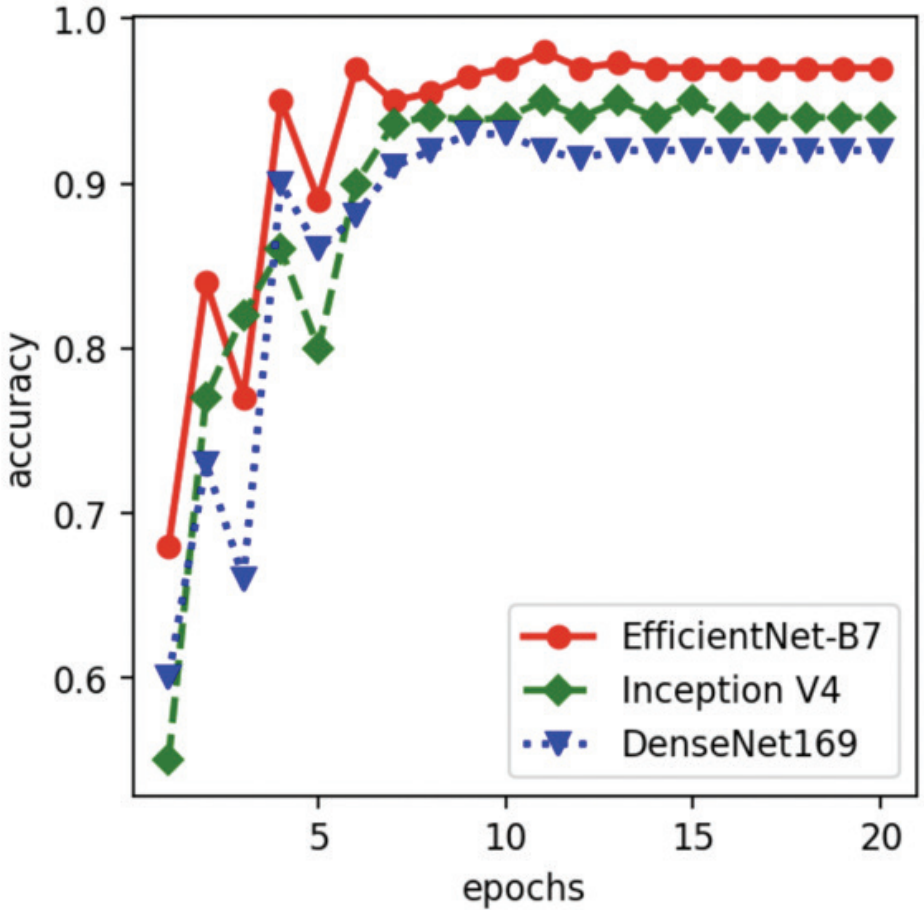

Figure 4. Accuracy convergence of three convolutional neural network models for ZEB2 expression image classification in cervical cancer.

Table 1. Comparison of accuracy, area under the curve (AUC), and F1 Score for convolutional neural network models in cervical cancer classification

Model	Accuracy	AUC	F1 Score
DenseNet169	0.92	0.93	0.92
Inception V4	0.94	0.96	0.95
EfficientNet-B7	0.97	0.98	0.97
Ensemble model	0.95	0.97	0.95

results, the EfficientNet-B7 model achieved the highest accuracy, AUC, and F1 score among the three models, with values of 0.97, 0.98, and 0.97, respectively. Inception V4 had the second-highest accuracy and AUC values of 0.94 and 0.96, respectively, and the highest F1 score of 0.95. DenseNet169 had the lowest values for all three metrics, with an accuracy of 0.92, AUC of 0.93, and F1 score of 0.92. The Ensemble model combining DenseNet169, Inception V4, and EfficientNet-B7 models achieved an accuracy of 0.95 and an AUC of 0.97, which was higher than the DenseNet169 and Inception V4 performance. The ROC curve of all models is shown in Figure 5. Overall, the evaluation results highlight the importance of choosing an appropriate CNN model for specific tasks. The EfficientNet-B7 model achieved the best performance in this study.

Finally, the Ensemble model was subjected to the Grad-CAM technique, which produced a visually informative heatmap that highlights the areas in the input image that played a crucial role in the classification outcome. Figure 6 shows the heatmap for a positive

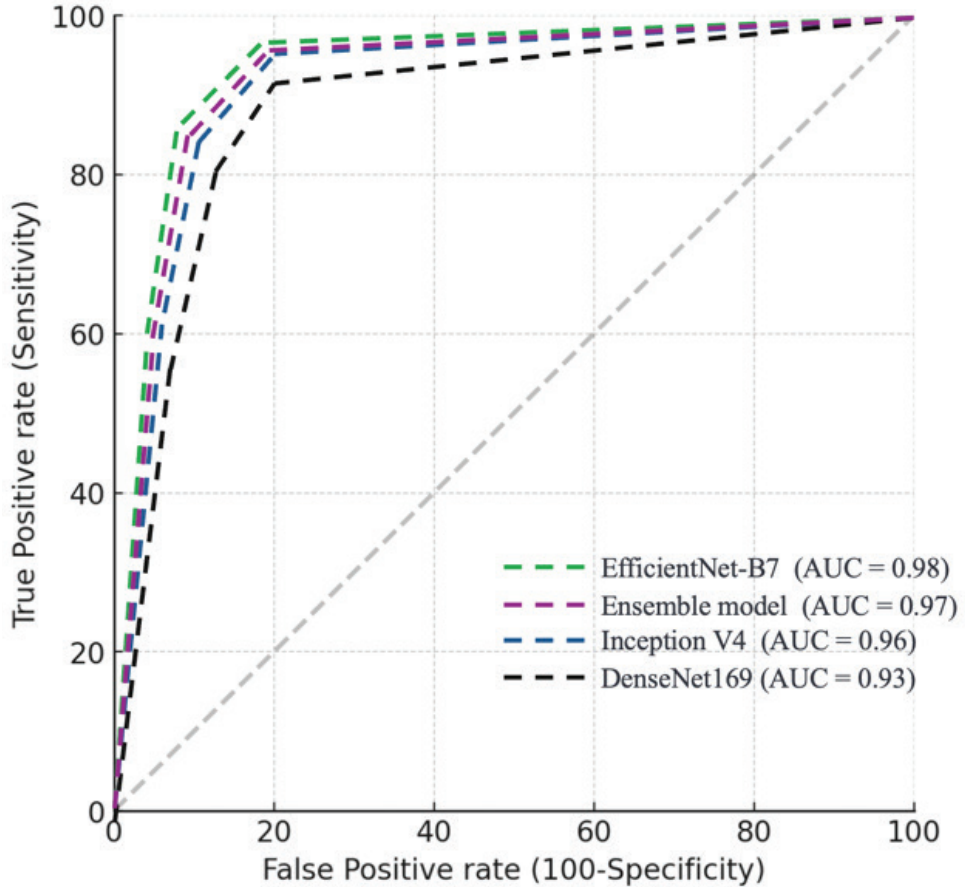

Figure 5. Receiver operating characteristic (ROC) curves demonstrating the diagnostic performance of convolutional neural network models for cervical cancer classification.

expression image of ZEB2 in CC. By observing the heatmaps generated by Grad-CAM, we can gain deeper insights into which areas of the image the model focuses on to make decisions. The hotspots were found in the upper-middle and lower-middle regions of the image. This result is consistent with a positive expression picture of CC by doctors' visual inspection. The utilisation of Grad-CAM visualisation enabled us to gain insights into the model's decision-making process by identifying the regions of the input image that significantly influenced the classification outcome. Employing Grad-CAM facilitated a comprehensive understanding of the hybrid model's capability in distinguishing between positive and negative ZEB2 expression images in CC.

Discussion

The results from the model training exercise provide compelling evidence that the EfficientNet-B7 model outperforms the other networks examined in this study. Achieving its peak performance after the twentieth epoch, it may owe its superior accuracy to its innovative compound scaling method (Tan and Le, 2019). By contrast, Inception V4 delivered the least accurate results, possibly owing to its relative complexity (Szegedy et al, 2017). DenseNet169, meanwhile, underperformed compared to the other models, yielding the least favourable outcomes across all performance metrics.

In-depth model evaluations further underscored the relative effectiveness of EfficientNet-B7, which outshone the other models across accuracy, AUC, and F1 score metrics. This suggests that it may represent the optimal model for classifying ZEB2 expression images in CC, offering considerable promise for future applications in this field. Yet despite its lower accuracy, Inception V4 posted strong results in terms of the F1 score, implying that it may be a suitable alternative for cases where identifying true positives is of paramount importance.

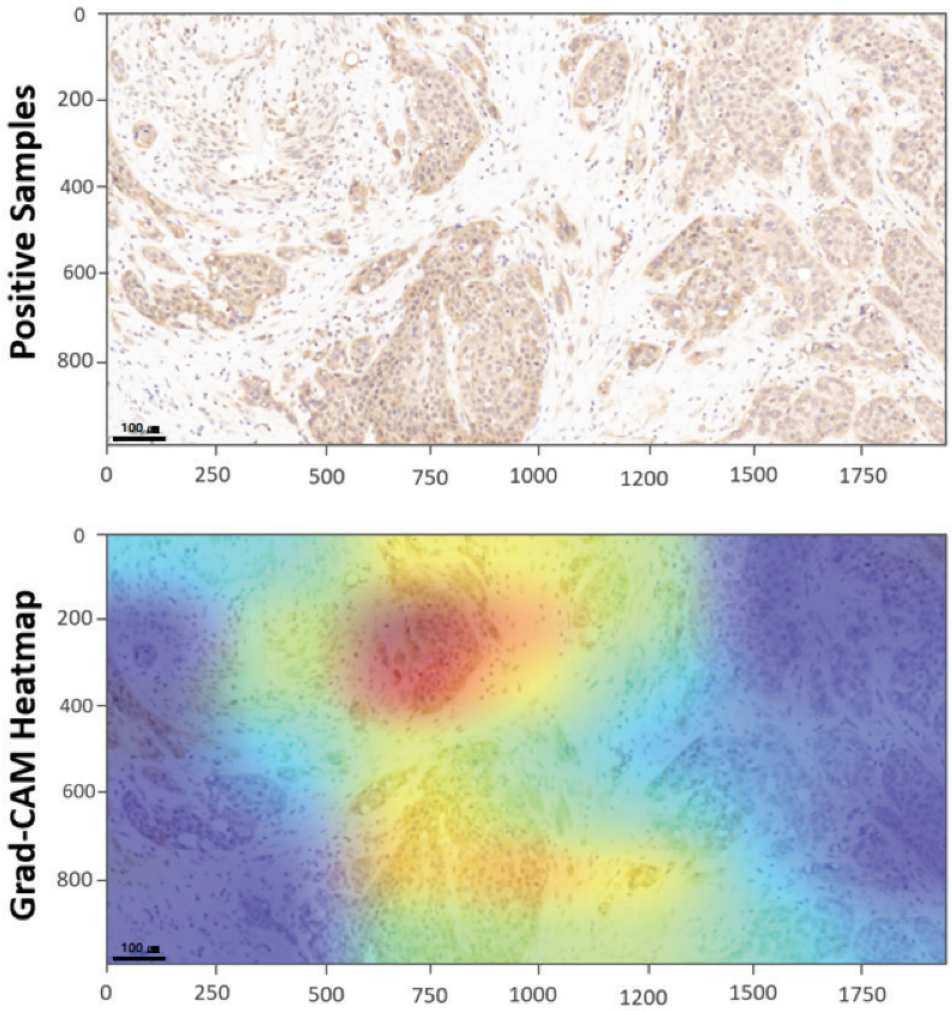

Figure 6. Gradient-weighted class activation mapping (Grad-CAM) heatmap visualisation of ZEB2 expression in cervical cancer using ensemble model. Scale bar=100 µm.

DenseNet169, on the other hand, demonstrated the lowest performance across all metrics, indicating that it may not be the most appropriate model for this specific task. However, the ensemble voting method—by combining the predictions from several models—showed a robust ability to reduce overfitting and enhance model generalizability (Kuncheva, 2014). This approach helps to mitigate the errors arising from individual models, ensuring reliable predictions even in instances where a single model underperforms.

A closer examination of EfficientNet-B7 reveals a uniquely balanced network, optimising depth, width, and resolution. This balance enables enhanced feature capture and pattern recognition in images while guarding against overfitting. Additionally, the model deploys a mix of efficient convolutional and depthwise convolutional layers, thereby reducing the computational requirements by lowering the parameter count (Tan and Le, 2019). Conversely, Inception V4 and DenseNet169, despite being longer-standing models, may not have been expressly optimised for the same level of efficiency and performance. Their unique architectural characteristics and strengths could potentially influence their performance in this task (Huang et al, 2017; Szegedy et al, 2017).

It is important to consider the broader contextual factors that can significantly influence model performance. Among these are the quality and volume of the dataset, the choice of hyperparameters during training, and the amount of available training data. Nonetheless, the findings of this study suggest that the EfficientNet-B7 model and the Ensemble model have significant potential in the accurate classification of ZEB2 expression images in CC, highlighting a promising avenue for future research and applications.

This study focused on classifying CC images based on ZEB2 expression. It is important to acknowledge that the results might be influenced by various external factors such as patient demographics, disease stage at the time of diagnosis, and specific medical treatments administered. While this research primarily analyzes the technical performance of the DL models, further studies could incorporate these variables to assess their impact on the accuracy and reliability of ZEB2 expression classification. For instance, the age and immune status of patients might modify the expression levels of ZEB2, while different treatment modalities such as chemotherapy or radiation therapy could alter cellular morphology, potentially affecting the model's predictions. Understanding these interactions is crucial for refining the model's application in clinical settings and ensuring robust performance across diverse patient populations.

While ZEB2 is recognised as a potential non-invasive biomarker for cancer diagnostics, its expression is not unique to CC. This raises a critical question about the specificity of the proposed DL models in distinguishing CC from other cancers that might also express ZEB2. To address this challenge, it would be beneficial to integrate multi-class classification strategies in future iterations of the model. Such strategies could involve training the model on a broader range of cancer types, allowing it to learn distinguishing features specific to each type. Additionally, incorporating clinical data such as patient history and coexisting conditions could improve the model's diagnostic accuracy and specificity.

However, there are two primary limitations to this research. Firstly, the dataset in the study is relatively small and imbalanced, which could impact the model's performance and generalizability. Secondly, the research solely concentrated on classifying ZEB2 expression images in CC, without taking into account other relevant clinical data which could also aid in accurate diagnosis and predicting disease progression such as patient age, HPV status, or histological grade. The additional exploration may be required to tackle these issues and amplify the precision and practicality of machine learning-driven CC diagnosis in the future.

Conclusion

In conclusion, this study proposes a CNN-based hybrid model for classifying ZEB2 expression images in CC using three different CNN models, namely EfficientNet, DenseNet, and InceptionNet. The approach uses an ensemble voting mechanism to merge forecasts from various models. Additionally, it incorporates Grad-CAM visualisation to emphasise the critical areas of the image that impacted the classification decision. The outcomes of the investigation indicate that this method attains exceptional precision when categorising CC ZEB2 expression images. Furthermore, it offers interpretability and transparency during the decision-making process. The proposed model has the potential to be a valuable tool for assisting medical professionals in diagnosing CC, ultimately improving patient outcomes and reducing healthcare costs.

Key points:

- Integrates multiple CNN models (EfficientNet, DenseNet, InceptionNet) using ensemble voting to increase robustness and accuracy in classifying ZEB2 expression in cervical cancer.
- Targets ZEB2, a key marker for tumour aggressiveness and progression in cervical cancer, enhancing the specificity and relevance of diagnostics.
- Employs Grad-CAM visualization to improve the interpretability of CNN decisions, providing visual explanations that highlight critical image regions influencing the classification.
- Achieves a high classification accuracy of 94.4%, demonstrating potential to improve early diagnosis and treatment outcomes in cervical cancer while reducing healthcare costs.

Author details

[1]Department of Gynecology, Women and Children's Hospital of Ningbo University, Ningbo, Zhejiang, China

Availability of data and materials

All data included in this study are available upon request by contact with the corresponding author.

Author contributions

CZ: Conceptualisation, methodology, writing—original draft; QS: Data curation, formal analysis; LZ: Investigation, data curation, methodology; YW: Conceptualisation, supervision, writing—review and editing. All authors contributed to important editorial changes in the manuscript. All authors have read and approved the final version of the manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics approval and consent to participate

This study was conducted in compliance with ethical standards and has been approved by the Ningbo Women and Children's Hospital ethics committee (approval number: 2023066). All participants signed an informed consent form.

Acknowledgement

Not applicable.

Funding

This research was supported by the Zhejiang Medicine and Health Science and Technology Project (grant no. 2020KY885 and 2020KY877). The funding bodies played no role in the design of the study, data collection, analysis, interpretation of data, or writing of the manuscript.

Conflict of interest

The authors declare there is no conflict of interest.

References

Al Husaini MAS, Habaebi MH, Gunawan TS et al. Thermal-based early breast cancer detection using inception V3, inception V4 and modified inception MV4. Neural Comput Appl. 2022;34(1):333–348. https://doi.org/10.1007/s00521-021-06372-1

Arezzo F, Cormio G, Loizzi V et al. HPV-negative cervical cancer: a narrative review. Diagnostics. 2021;11(6):952. https://doi.org/10.3390/diagnostics11060952

Bray F, Ferlay J, Soerjomataram I et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424. https://doi.org/10.3322/caac.21492

Ceylan Z, Pekel E. Comparison of multi-label classification methods for prediagnosis of cervical cancer. Graph Models. 2017;21:22. https://doi.org/10.18201/ijisae.82426

Chandran V, Sumithra MG, Karthick A et al. Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images. Biomed Res Int. 2021;2021:5584004. https://doi.org/10.1155/2021/5584004

Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet. 2019;393(10167):169–182. https://doi.org/10.1016/S0140-6736(18)32470-X

Cox S, American Society for Colposcopy and Cervical Pathology, American College of Obstetricians and Gynecologists. Guidelines for papanicolaou test screening and follow-up. J Midwifery Womens Health. 2012;57(1):86–89. https://doi.org/10.1111/j.1542-2011.2011.00116.x

Demarco M, Hyun N, Carter-Pokras O et al. A study of type-specific HPV natural history and implications for contemporary cervical cancer screening programs. EClinicalMedicine. 2020;22:100293. https://doi.org/10.1016/j.eclinm.2020.100293

Diao P, Ge H, Song Y et al. Overexpression of ZEB2-AS1 promotes epithelial-to-mesenchymal transition and metastasis by stabilizing ZEB2 mRNA in head neck squamous cell carcinoma. J Cell Mol Med. 2019;23(6):4269–4280. https://doi.org/10.1111/jcmm.14318

Eddy DM. Secondary prevention of cancer: an overview. Bull World Health Organ. 1986;64(3):421–429.

Feng S, Liu W, Bai X et al. LncRNA-CTS promotes metastasis and epithelial-to-mesenchymal transition through regulating miR-505/ZEB2 axis in cervical cancer. Cancer Lett. 2019;465:105–117. https://doi.org/10.1016/j.canlet.2019.09.002

Gibb RK, Martens MG. The impact of liquid-based cytology in decreasing the incidence of cervical cancer. Rev Obstet Gynecol. 2011;4(Suppl 1):S2–S11. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3101960/

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Paper presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21–26, 2017; Honolulu, HI, USA. Accessed July 2, 2024. https://doi.org/10.1109/CVPR.2017.243

Hull R, Mbele M, Makhafola T et al. Cervical cancer in low and middle-income countries. Oncol Lett. 2020;20(3):2058–2074. https://doi.org/10.3892/ol.2020.11754

Iandola F, Moskewicz M, Karayev S et al. DenseNet: Implementing efficient convnet descriptor pyramids. 2014. https://doi.org/10.48550/arXiv.1404.1869

Jusman Y, Ng SC, Abu Osman NA. Intelligent screening systems for cervical cancer. Sci World J. 2014;2014:1–15. https://doi.org/10.1155/2014/810368

Kang Z, Li Y, Liu J et al. H-CNN combined with tissue Raman spectroscopy for cervical cancer detection. Spectrochim Acta A Mol Biomol Spectrosc. 2023;291:122339. https://doi.org/10.1016/j.saa.2023.122339

Kannojia SP, Jaiswal G. Ensemble of hybrid CNN-ELM model for image classification. Paper presented at: 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN); February 22–23, 2018; Noida, India. Accessed July 2, 2024. https://doi.org/10.1109/SPIN.2018.8474196

Kessler TA. Cervical cancer: prevention and early detection. Semin Oncol Nurs. 2017;33(2):172–183. https://doi.org/10.1016/j.soncn.2017.02.005

Khamparia A, Gupta D, de Albuquerque VHC, Sangaiah AK, Jhaveri RH. Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning. J Supercomput. 2020;76(11):8590–8608. https://doi.org/10.1007/s11227-020-03159-4

Khan IU, Aslam N, Anwar T et al. Remote diagnosis and triaging model for skin cancer using EfficientNet and extreme gradient boosting. Complexity. 2021;2021:1–13. https://doi.org/10.1155/2021/5591614

Ko H, Ha H, Cho H, Seo K, Lee J. Pneumonia detection with weighted voting ensemble of CNN models. Paper presented at: 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD); May 25–28, 2019; Chengdu, China. Accessed July 2, 2024. https://doi.org/10.1109/ICAIBD.2019.8837042

Kuncheva LI. Combining pattern classifiers: methods and algorithms. 1st edn. John Wiley and Sons: Hoboken, NJ, USA. 2014.

Li C, Chen H, Li X et al. A review for cervical histopathology image analysis using machine vision approaches. Artif Intell Rev. 2020;53(7):4821–4862. https://doi.org/10.1007/s10462-020-09808-7

Li MZ, Wang JJ, Yang SB et al. ZEB2 promotes tumor metastasis and correlates with poor prognosis of human colorectal cancer. Am J Transl Res. 2017;9(6):2838–2851. https://pubmed.ncbi.nlm.nih.gov/28670373/

Rayavarapu K, Krishna KKV. Prediction of cervical cancer using voting and DNN classifiers. Paper presented at: 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT); March 01–03, 2018; Coimbatore, India. Accessed July 2, 2024. https://doi.org/10.1109/ICCTCT.2018.8551176

Selvaraju RR, Cogswell M, Das A et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Paper presented at: Proceedings of the IEEE International Conference on Computer Vision; October 22–29, 2017; Venice, Italy. Accessed July 2, 2024. https://doi.org/10.48550/arXiv.1610.02391

Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6(1):1–48. https://doi.org/10.1186/s40537-019-0197-0

Singh M, Spoelstra NS, Jean A et al. ZEB1 expression in type I vs type II endometrial cancers: a marker of aggressive disease. Modern Pathol. 2008;21(7):912–923. https://doi.org/10.1038/modpathol.2008.82

Small W, Bacon MA, Bajaj A et al. Cervical cancer: a global health crisis. Cancer. 2017;123(13):2404–2412. https://doi.org/10.1002/cncr.30667

Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resNet and the impact of residual connections on learning. Paper presented at: Proceedings of the AAAI Conference on Artificial Intelligence; February 4–9, 2017; San Francisco, CA, USA. Accessed July 2, 2024. https://doi.org/10.1609/aaai.v31i1.11231

Taha B, Dias J, Werghi N. Classification of cervical-cancer using pap-smear images: A convolutional neural network approach. Paper presented at: Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017; July 11–13, 2017; Edinburgh, UK. Accessed July 2, 2024. https://doi.org/10.1007/978-3-319-60964-5_23

Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. Paper presented at: International Conference on Machine Learning; June 9–15, 2019; Long Beach, CA, USA. Accessed July 2, 2024. https://doi.org/10.48550/arXiv.1905.11946

Tian R, Cui Z, He D et al. Risk stratification of cervical lesions using capture sequencing and machine learning method based on HPV and human integrated genomic profiles. Carcinogenesis. 2019;40(10):1220–1228. https://doi.org/10.1093/carcin/bgz094

Toratani M, Konno M, Asai A et al. A convolutional neural network uses microscopic images to differentiate between mouse and human cell lines and their radioresistant clones. Cancer Res. 2018;78(23):6703–6707. https://doi.org/10.1158/0008-5472.CAN-18-0653

Tseng CJ, Lu CJ, Chang CC, Chen GD. Application of machine learning to predict the recurrence-proneness for cervical cancer. Neural Comput Appl. 2014;24:1311–1316. https://doi.org/10.1007/s00521-013-1359-1

William W, Ware A, Basaza-Ejiri AH, Obungoloch J. A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. Comput Methods Programs Biomed. 2018;164:15–22. https://doi.org/10.1016/j.cmpb.2018.05.034

Xue P, Wang J, Qin D et al. Deep learning in image-based breast and cervical cancer detection: a systematic review and meta-analysis. NPJ Digit Med. 2022;5(1):19. https://doi.org/10.1038/s41746-022-00559-z

Ye C, Hu Y, Wang J. MicroRNA-377 targets zinc finger E-box-binding homeobox 2 to inhibit cell proliferation and invasion of cervical cancer. Oncol Res. 2019;27(2):183–192. https://doi.org/10.3727/096504018X15201124340860

Zhang J, Liu Y. Cervical cancer detection using SVM based feature screening. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; September 26–29, 2004; Saint-Malo, France. Accessed July 2, 2024. https://doi.org/10.1007/978-3-540-30136-3_106

Zhong Z, Zheng M, Mai H, Zhao J, Liu X. Cancer image classification based on DenseNet model. J Phys, Conf Ser. 2020;1651(1):012143. https://doi.org/10.1088/1742-6596/1651/1/012143