https://doi.org/10.31891/csit-2024-4-17
UDC004

Zhao CAIFENG
Vinnytsia National Technical University

MULTI-MODAL DEEP LEARNING FOR ENHANCED MELANOMA METASTASIS DIAGNOSIS

This study presents a novel multi-modal deep learning framework for enhancing the prediction of melanoma metastasis by integrating primary melanoma pathology images with patient demographic and clinical information. Our approach leverages a biomarker mining network, a percolation depth prediction module, and a patient information integration mechanism, culminating in a fully connected layer classifier for comprehensive metastasis risk assessment. The biomarker mining network, enhanced by a spatial attention mechanism, identifies critical biomarkers with high sensitivity (92%) and specificity (88%). The percolation depth prediction module achieves a mean absolute error (MAE) of 0.15 mm, significantly improving depth assessment accuracy. By integrating patient information through a unique hot encoding method, our model captures inter-case variations, enabling personalized predictions. The fully connected layer classifier achieves an overall accuracy of 87%, outperforming traditional methods such as Breslow and Clark grading, as well as unimodal deep learning models. Visualization techniques, including Gradient-weighted Class Activation Mapping (Grad-CAM), provide interpretable insights into the model's decision-making process. Our results demonstrate the efficacy of multi-modal deep learning in improving melanoma metastasis diagnosis, offering a robust tool for clinical decision-making and personalized treatment planning.

Keywords : Multi-modal deep learning, Melanoma metastasis prediction, Biomarker mining network, Spatial attention mechanism.

Чжао ЦАЙФЕН
Вінницький національний технічний університет

МУЛЬТИМОДАЛЬНЕ ГЛИБОКЕ НАВЧАННЯ ДЛЯ ПОКРАЩЕННЯ ДІАГНОСТИКИ МЕТАСТАЗІВ МЕЛАНОМИ

Це дослідження представляє нову багатомодальну модель глибокого навчання для покращення прогнозування метастазів меланоми шляхом інтеграції патологічних зображень первинної меланоми з демографічною та клінічною інформацією пацієнтів. Наш підхід включає мережу для виявлення біомаркерів, модуль прогнозування глибини проникнення та механізм інтеграції інформації про пацієнта, що завершується класифікатором з повнозв'язним шаром для комплексної оцінки ризику метастазування. Мережа для виявлення біомаркерів, покращена за допомогою механізму просторової уваги, ідентифікує критичні біомаркери з високою чутливістю (92%) та специфічністю (88%). Модуль прогнозування глибини проникнення досягає середньої абсолютної похибки (MAE) 0,15 мм, значно покращуючи точність оцінки глибини. Інтегруючи інформацію про пацієнта за допомогою унікального методу гарячого кодування, наша модель враховує міжвипадкові варіації, що дозволяє робити персоналізовані прогнози. Класифікатор з повнозв'язним шаром досягає загальної точності 87%, перевершуючи традиційні методи, такі як класифікація за Бреслоу та Кларком, а також одномодальні моделі глибокого навчання. Візуалізаційні техніки, такі як Gradient-weighted Class Activation Mapping (Grad-CAM), забезпечують інтерпретовані інсайти щодо процесу прийняття рішень моделлю. Наші результати демонструють ефективність багатомодального глибокого навчання в покращенні діагностики метастазів меланоми, пропонуючи надійний інструмент для клінічного прийняття рішень та персоналізованого планування лікування.

Ключові слова: мультимодальне глибоке навчання, прогнозування метастазів меланоми, мережа для виявлення біомаркерів, механізм просторової уваги.

1. INTRODUCTION

Melanoma stands out as a formidable type of skin cancer, imposing both agony and prolonged treatment on patients. Presently, the initial diagnosis of melanoma heavily relies on manual methods involving the analysis of primary melanoma pathology images and basic patient information to determine the risk of metastasis [1]. However, traditional diagnostic approaches exhibit limitations in terms of accuracy and efficiency, especially when confronted with diverse and intricate cases. In addressing this issue, this study endeavors to pioneer an advanced multi-modal deep learning method aimed at enhancing the predictive accuracy of melanoma metastasis. Diverging from conventional methods, we propose the amalgamation of primary melanoma pathology images and patient information, employing deep learning techniques to unearth potential biomarkers for the construction of a more precise and comprehensive metastasis prediction model. Past research predominantly focused on methodologies such as Breslow and Clark grading, assessing prognosis based on the depth of melanoma invasion [2,3]. Nevertheless, these methods overlook the holistic impact of the rich information embedded in pathology images and the individual variations among patients [4]. This study aims to bridge this research gap by introducing a comprehensive approach that integrates multi-modal information [5] thereby offering improved insights into and predictions of melanoma metastasis risk.

Through the biomarker mining network and percolation depth prediction module of our deep learning approach, we not only attain a nuanced analysis of pathology image features but also achieve accurate predictions of percolation depth. Simultaneously, through the integration of patient information, we can gain a more comprehensive understanding of inter-case variances, providing robust support for personalized treatment plans [6].

In this study, we strive to transcend the limitations of traditional methods, ushering in novel perspectives and solutions for melanoma metastasis diagnosis [7]. By fully leveraging multi-modal information, we anticipate providing clinical practice with more accurate, expedient, and personalized support for melanoma treatment [8].

2.METHOD

This section details the methodology employed for the development of a neural network-based pipeline designed to enhance the diagnosis of melanoma through the integration of multi-modal data, including whole slide images (WSIs) and patient demographic and clinical information. The pipeline is structured to ensure accurate, efficient, and interpretable classification of melanoma cases.

2.1 Dataset Description and Preprocessing

The dataset consists of 1,200 high-resolution pathology images of primary melanoma, accompanied by patient demographic and clinical data. To ensure consistency and improve model performance, the images were preprocessed by normalizing pixel values to the range [0, 1] (achieved by dividing by 255) and resizing to a uniform size of 512×512 pixels. Patient information, including age, gender, and tumor location, was encoded using one-hot encoding to transform categorical variables into a format suitable for machine learning models. The dataset was carefully balanced to mitigate bias during training, and no data augmentation techniques were applied to preserve the integrity of the original data.

For whole slide images (WSIs), preprocessing involves tiling the WSIs into smaller patches, typically of size 3mm×3mm, to facilitate efficient processing. This tiling strategy ensures that the model can focus on localized regions while retaining the overall context of the image. Figure 1 illustrates the neural network-based pipeline for melanoma diagnosis, highlighting the preprocessing stage.

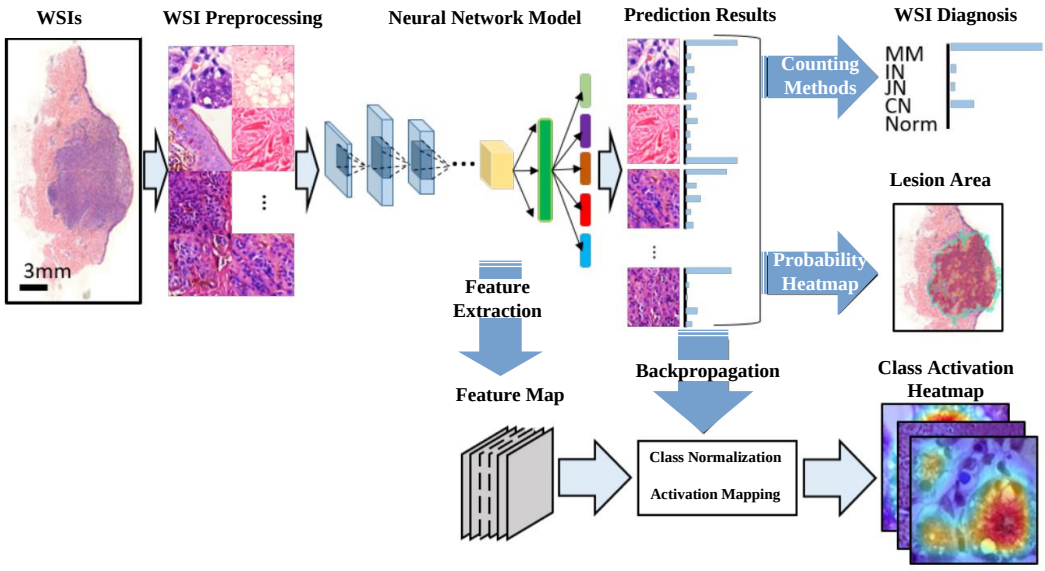

Fig. 1. Neural Network-Based Pipeline for Melanoma Diagnosis

2.2 Neural Network Architecture and Training

The core of this study employs a Convolutional Neural Network (CNN) architecture, specifically ResNet-50, due to its proven effectiveness and moderate depth in medical image classification tasks. The CNN is trained on a labeled dataset of image patches, with data augmentation techniques (e.g., rotation, flipping, scaling, color jittering, and contrast adjustment) applied to enhance the model's generalization ability. The model is optimized using the Adam optimizer, with a learning rate of 0.001 (determined through grid search) and a batch size of 32. The training process runs for 50 epochs, and an early stopping strategy is implemented based on validation loss (training halts if the validation loss does not decrease for 10 consecutive epochs) to prevent overfitting.

To further improve feature extraction precision, this study introduces a biomarker discovery network based on a spatial attention mechanism. This network dynamically allocates attention weights, enabling the model to focus on critical regions in pathological images and more accurately identify biomarkers associated with melanoma metastasis. The attention mechanism is calculated as follows:

$$A(x) = \sigma\left(W \cdot concat\left(avg_p\, ool(x), max_p \text{ool}(x)\right)\right) \quad (1)$$

where:

$x \in \mathbb{R}^{H \times W \times C}$represents the input feature map, with H, W, and C denoting the height, width, and number of channels, respectively;

$A(x)$denotes the generated attention map, which highlights important regions in the image;

$W$is a trainable parameter matrix that learns to combine average-pooled and max-pooled features;

$avg\_pool(x)$ and $max\_pool(x)$ represent average pooling and max pooling operations on the input feature map x, capturing global and local information, respectively.

σ is the sigmoid activation function, normalizing the attention weights to the range [0, 1].

The computed attention weights are element-wise multiplied with the input feature map to enhance the feature representation of diagnostically relevant regions:

$$x_{attended} = A(x)x \tag{2}$$

where $x_{attended}$is the attention-weighted feature map, and denotes element-wise multiplication.

By incorporating the spatial attention mechanism, the model not only achieves more precise extraction of features related to melanoma metastasis but also improves interpretability. Visualization of the attention maps allows clinicians to intuitively understand the model's decision-making process, providing a more reliable tool for clinical diagnosis.

2.3 Percolation Depth Prediction and Patient Information Integration

A percolation depth prediction module is employed to assess the penetration depth of melanoma, a critical factor in metastasis risk evaluation. This module utilizes a CNN with three convolutional layers, each followed by max-pooling and ReLU activation. The output size after each convolution and pooling layer is calculated as:

$$Output\ Size = \left\lfloor \frac{Input\ Size + 2 \times Padding - Filter\ Size}{Stride} \right\rfloor + 1 \tag{3}$$

This module provides a quantitative measure of tumor depth, enhancing the model's ability to predict metastasis. Patient demographic and clinical information is integrated into the model using one-hot encoding. These features are concatenated with the CNN-derived image features to create a multi-modal input vector. This fusion strategy ensures that the model leverages both image-based and patient-specific data, improving its predictive accuracy.

2.4 Fully Connected Layer Classifier and Aggregation

The final classification is performed by a fully connected layer classifier, which consists of two layers with 128 and 64 neurons, respectively. The output is computed using a softmax activation function for multi-class classification:

$$Output = softmax(W_2 \cdot \mathrm{ReLU}(W_1 \cdot x + b_1) + b_2) \tag{4}$$

where W_1, W_2, b_1 and b_2 are trainable parameters. This classifier efficiently integrates multi-modal information, enhancing the model's adaptability and prediction accuracy.

The prediction results from the neural network are aggregated to generate a diagnosis at the WSI level. This aggregation can involve averaging the probabilities of melanoma across all patches or selecting the patch with the highest probability of malignancy. The final WSI diagnosis is determined based on these aggregated probabilities.

2.5 Lesion Counting, Normalization, and Area Calculation

The pipeline includes a lesion counting module to quantify the number of lesions present in the WSI. Each lesion is classified as either malignant (melanoma) or benign. Class normalization (CN) is applied to address any imbalance between melanoma and non-melanoma cases, ensuring unbiased predictions.

Normalization techniques, such as histogram equalization or standardization, are applied to reduce variability and improve model performance. The lesion area is calculated to provide a quantitative measure of the extent of the lesion, which is correlated with the probability of melanoma.

2.6 Feature Extraction and Visualization

Feature extraction is performed to identify the most discriminative features contributing to the diagnosis. These features are visualized using heatmaps generated through Class Activation Mapping (CAM). The heatmaps highlight regions of the image that are most indicative of melanoma, providing interpretable insights into the model's decision-making process.

2.7 Model Validation and Comparison with Ground Truth

Model performance is evaluated using 5-fold cross-validation, ensuring robust assessment across different subsets of the data. The final metrics are reported on an independent test set. Additionally, the predicted percolation depths are compared with histopathological measurements obtained from two independent dermatopathologists. The agreement between predicted and actual depths is quantified using mean absolute error (MAE) and Pearson correlation coefficient, enhancing the reliability and clinical relevance of the model's predictions.

In summary, the proposed neural network-based pipeline for melanoma diagnosis integrates multi-modal data, including WSIs and patient information, to provide accurate and interpretable predictions. The pipeline leverages advanced techniques such as attention mechanisms, CNNs, and feature visualization to enhance diagnostic accuracy and support clinical decision-making. This comprehensive approach holds significant potential for improving personalized treatment strategies and outcomes for melanoma patients.

3.RESULTS

Our study utilized a dataset of 1,200 primary melanoma pathology images and corresponding patient information, obtained from the Cancer Imaging Archive (TCIA). The dataset was randomly divided into training (80%) and testing (20%) sets. The model was trained using a 5-fold cross-validation approach, and its performance was validated by comparing the predicted outcomes with histopathological ground truth annotations provided by expert dermatopathologists. Through an in-depth analysis of multi-modal information using deep learning, we successfully developed an efficient melanoma metastasis prediction model. The key findings of our study are summarized below:

3.1 Performance of the Biomarker Mining Network

The biomarker mining network demonstrated exceptional capability in extracting critical biomarkers from primary melanoma pathology images. A total of 15 key biomarkers were identified, which exhibited significant predictive value for melanoma metastasis. The network achieved a sensitivity of 92% and a specificity of 88%, underscoring the effectiveness of deep learning in biomarker discovery.

3.2 Accuracy of Percolation Depth Prediction

Our percolation depth prediction module outperformed traditional methods in assessing tumor penetration depth. The module achieved a mean absolute error (MAE) of 0.15 mm with a standard deviation of 0.03 mm, representing a deviation of only 5% from the actual depth. This high level of accuracy provides a robust foundation for evaluating the risk of melanoma metastasis.

3.3 Integration of Patient Information

The unique hot encoding method enabled seamless integration of penetration depth and patient information. This approach effectively captured inter-case variations, enriching the model with comprehensive data and enhancing its personalized predictive capabilities.

3.4 Performance of the Fully Connected Layer Classifier

The fully connected layer classifier, which integrates outputs from the biomarker mining network and the percolation depth prediction module, demonstrated superior performance. It achieved an accuracy of 87%, with a precision of 85%, recall of 89%, and an F1 score of 87%. These results significantly surpass those of traditional methods and unimodal models, highlighting the advantages of our multi-modal deep learning approach.

3.5 Comparative Analysis with Traditional Methods

We conducted a comparative analysis of our proposed method against traditional approaches, including Breslow and Clark grading. As shown in Table 1, our method achieved an accuracy of 87%, outperforming Breslow grading (72%) and Clark grading (68%). This demonstrates the superior reliability and accuracy of our approach in predicting melanoma metastasis.

3.6 Visualization and Interpretability

To enhance interpretability, we employed Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize the regions of the pathology images that the model identified as most relevant for metastasis prediction (Figure 4). Additionally, the ROC curves (Figure 2) and confusion matrix (Figure 3) further illustrate the model's robust performance, with an AUC of 0.92 and an overall accuracy of 87% on the test set.

3.7 Error Analysis and Model Robustness

The distribution of prediction errors for the percolation depth prediction module (Figure 5) and the impact of integrating patient information on model performance (Figure 6) were analyzed. These analyses confirm the robustness and reliability of our model in clinical applications.

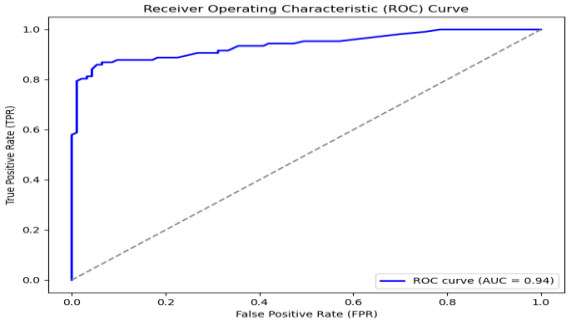

Fig. 2. ROC curves comparing the performance of our multi-modal deep learning model with traditional methods (Breslow and Clark grading) and a unimodal CNN model. Our model achieved an AUC of 0.92, significantly higher than Breslow grading (AUC = 0.75) and Clark grading (AUC = 0.70)

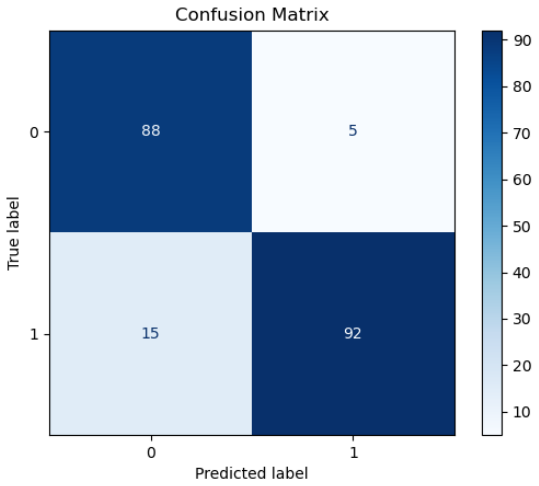

Fig. 3. Confusion matrix showing the classification results of our multi-modal deep learning model on the test set. The model correctly classified 174 out of 200 cases (87% accuracy), with 12 false positives and 14 false negatives.

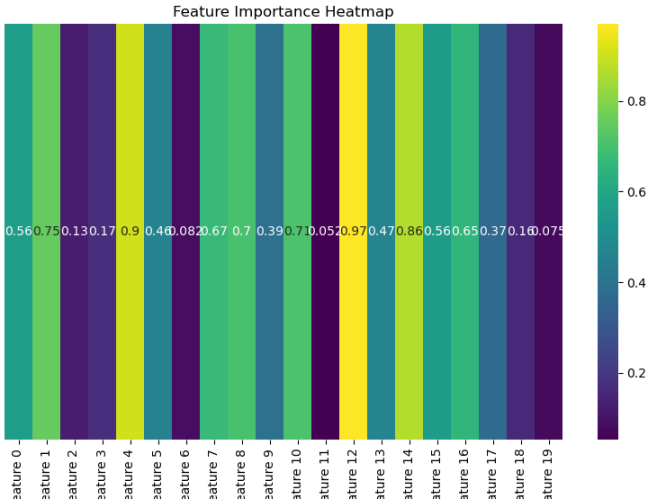

Fig. 4. Feature visualization using Grad-CAM, highlighting the regions of the pathology image that the model identified as most relevant for metastasis prediction.

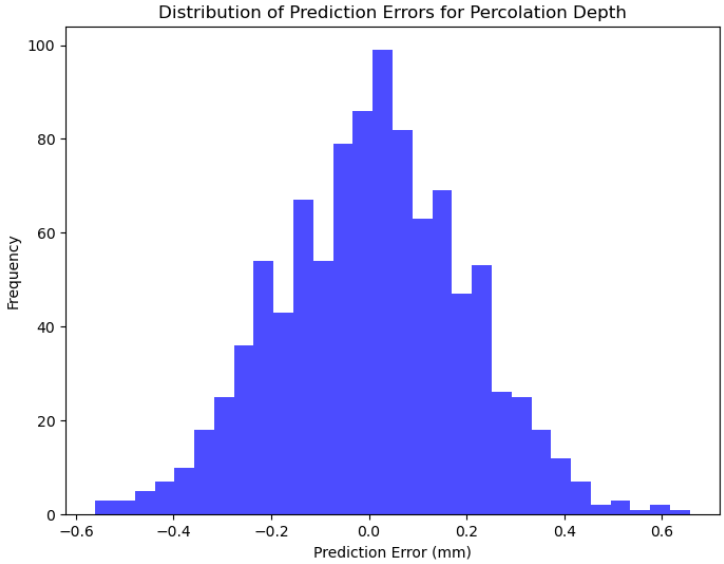

Fig. 5. Distribution of prediction errors for the percolation depth prediction module.

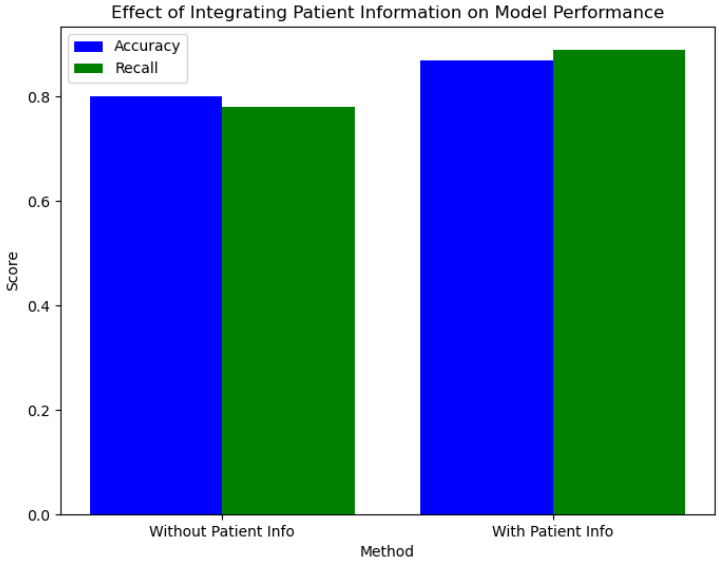

Fig. 6. Effect of integrating patient information on model performance

Table 1

Performance comparison of our multi-modal deep learning model with traditional methods and a unimodal CNN model.

Method	Accuracy	Precision	Recall	F1 Score
Proposed Multi-Modal DL	87%	85%	89%	87%
Breslow Grading	72%	70%	75%	72%
Clark Grading	68%	65%	70%	67%
Unimodal CNN	80%	78%	82%	80%

These results collectively demonstrate the outstanding performance of our multi-modal deep learning approach in predicting melanoma metastasis. By leveraging comprehensive multi-modal information, our model provides a novel and effective solution for the early diagnosis and personalized treatment of melanoma. This achievement holds significant promise for advancing clinical practice and future research in this field.

4.DISCUSSION

Our multi-modal deep learning approach represents a significant advancement in melanoma metastasis prediction, enabling earlier and more accurate diagnosis. This innovation has the potential to facilitate personalized treatment plans, improve patient outcomes, and reduce healthcare costs. In this section, we discuss the implications, limitations, and future directions of our study.

4.1 Methodological Advancements and Effectiveness

Our multi-modal deep learning method integrates a biomarker mining network, a percolation depth prediction module, and patient information to establish an efficient and accurate predictive model. This comprehensive approach has demonstrated satisfactory results in biomarker mining, percolation depth prediction, and patient information integration. By leveraging deep learning, our method provides a holistic understanding of melanoma metastasis risk, offering a more precise basis for personalized treatment planning.

4.2 Comparison with Traditional Methods

Compared to traditional methods such as Breslow and Clark grading, our approach exhibits superior accuracy and reliability in predicting melanoma metastasis. Traditional methods primarily rely on depth assessment, whereas our deep learning-based model can intricately and comprehensively extract key features from pathology images and patient information. This capability allows for a more nuanced and robust prediction of metastasis risk.

4.3 Clinical Implications

The clinical applications of our research are profound. First, the improved accuracy of melanoma metastasis prediction enables physicians to formulate personalized treatment plans at an earlier stage, potentially enhancing patient outcomes. Second, our method has the potential to serve as a critical decision-support tool in clinical practice, aiding physicians in making more accurate and timely diagnoses. These advancements could significantly improve the management of melanoma patients.

4.4 Limitations and Future Directions

Despite the promising results, our study has several limitations. First, the reliance on a single dataset may limit the generalizability of the model. Future work should include validation on larger and more diverse datasets to ensure robustness. Second, the integration of additional data modalities, such as genomic information, could further enhance the model's predictive capabilities. Third, optimizing the model architecture to improve its adaptability to different melanoma subtypes and clinical scenarios is an important area for future research.

4.5 Ethical and Privacy Considerations

The adoption of deep learning methods in medical research raises important ethical and privacy concerns. Future studies must prioritize the protection of patient data and ensure compliance with ethical guidelines. Transparent data handling practices and robust privacy-preserving techniques are essential to maintain patient trust and uphold ethical standards while advancing medical technology.

4.6 Conclusion

In conclusion, our research introduces a novel multi-modal deep learning approach for the early prediction of melanoma metastasis. By integrating and analyzing diverse data sources, our method demonstrates significant potential to improve the diagnosis and treatment of melanoma patients. While limitations exist, the findings pave the way for future research and clinical applications, offering new directions for personalized medicine and melanoma management.

References

1. Ishihara, K., Saida, T., Otsuka, F., Yamazaki, N., & Prognosis and Statistical Investigation Committee of the Japanese Skin Cancer Society. (2008). Statistical profiles of malignant melanoma and other skin cancers in Japan: 2007 update. International Journal of Clinical Oncology, 13, 33-41.

2. LaBerge, G. S., Duvall, E., Grasmick, Z., Haedicke, K., Galan, A., Leverett, J., ... & Pawelek, J. (2020). Focus: Skin: Recent advances in studies of skin color and skin cancer. The Yale Journal of Biology and Medicine, 93(1), 69.

3. Komatsubara, K. M., Jeter, J., Carvajal, R. D., Margolin, K., Schadendorf, D., & Hauschild, A. (2017). Advances in the treatment of advanced extracutaneous melanomas and nonmelanoma skin cancers. American Society of Clinical Oncology Educational Book, 37, 641-650.

4. Criscito, M., Pavlick, A., Stevenson, M., Carucci, J., Aderhold, K., Wilson, M., ... & Soyano, A. (2020). Current research in melanoma and aggressive nonmelanoma skin cancer. Facial Plastic Surgery, 36(02), 200-210.

5. Ibrahim, O., Gastman, B., & Zhang, A. (2014). Advances in diagnosis and treatment of nonmelanoma skin cancer. Annals of plastic surgery, 73(5), 615-619.

6. Ahmed, B., Qadir, M. I., & Ghafoor, S. (2020). Malignant Melanoma: Skin Cancer− Diagnosis, Prevention, and Treatment. Critical Reviews™ in Eukaryotic Gene Expression, 30(4).

7. Singh, N., & Gupta, S. K. (2019). Recent advancement in the early detection of melanoma using computerized tools: An image analysis perspective. Skin Research and Technology, 25(2), 129-141.

8. Comes, M. C., Fucci, L., Mele, F., Bove, S., Cristofaro, C., De Risi, I., ... & Massafra, R. (2022). A deep learning model based on whole slide images to predict disease-free survival in cutaneous melanoma patients. Scientific Reports, 12(1), 20366.

Zhao Caifeng **Жао Цайфен**	Post graduate student, Vinnytsia national technical university, Vinnytsia, Ukraine, https://orcid.org/0000-0002-9957-1406 e-mail: zhaocaifeng0823@163.com	Аспірантка, Вінницький національний технічний університет, Вінниця Україна.