Virtual multiplexed immunofluorescence staining from non-antibody-stained fluorescence imaging for gastric cancer prognosis

Zixia Zhou,[a,*,e] Yuming Jiang,[b,**,e] Zepang Sun,[c] Taojun Zhang,[c] Wanying Feng,[d] Guoxin Li,[c] Ruijiang Li,[a] and Lei Xing[a,***]

[a]Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA, 94305, USA
[b]Department of Radiation Oncology, Wake Forest University School of Medicine, Winston Salem, NC, 27109, USA
[c]Department of General Surgery & Guangdong Provincial Key Laboratory of Precision Medicine for Gastrointestinal Tumor, Nanfang Hospital, Southern Medical University, 510515, Guangzhou, China
[d]Department of Pathology, School of Basic Medical Sciences, Southern Medical University, 510515, Guangzhou, China

eBioMedicine
2024;107: 105287
Published Online xxx
https://doi.org/10.1016/j.ebiom.2024.105287

Summary

Background Multiplexed immunofluorescence (mIF) staining, such as CODEX and MIBI, holds significant clinical value for various fields, such as disease diagnosis, biological research, and drug development. However, these techniques are often hindered by high time and cost requirements.

Methods Here we present a Multimodal-Attention-based virtual mIF Staining (MAS) system that utilises a deep learning model to extract potential antibody-related features from dual-modal non-antibody-stained fluorescence imaging, specifically autofluorescence (AF) and DAPI imaging. The MAS system simultaneously generates predictions of mIF with multiple survival-associated biomarkers in gastric cancer using self- and multi-attention learning mechanisms.

Findings Experimental results with 180 pathological slides from 94 patients with gastric cancer demonstrate the efficiency and consistent performance of the MAS system in both cancer and noncancer gastric tissues. Furthermore, we showcase the prognostic accuracy of the virtual mIF images of seven gastric cancer related biomarkers, including CD3, CD20, FOXP3, PD1, CD8, CD163, and PD-L1, which is comparable to those obtained from the standard mIF staining.

Interpretation The MAS system rapidly generates reliable multiplexed staining, greatly reducing the cost of mIF and improving clinical workflow.

Funding Stanford 2022 HAI Seed Grant; National Institutes of Health 1R01CA256890.

Keywords: Multiplexed immunofluorescence staining; Artificial intelligence; Gastric cancer prognostic; Immune cell biomarker

Introduction

Pathological functional staining by multiplexed immunofluorescence (mIF) allows accurate identification and diagnosis of various diseases, ranging from cancer and infectious diseases to autoimmune disorders and genetic disorders,[1–3] and provides an indispensable tool for individualised patient care. By visualising the presence and precise location of specific biomolecules in tissue samples, mIF empowers pathologists with crucial insights into underlying biology and facilitates informed treatment decisions,[4] especially in prognostic prediction and selection of patients for immunotherapy.[5] While multiplexing is highly desirable for the detection and location of diverse antigens in various tissues,[6] existing techniques, such as multiplexed ion beam imaging (MIBI)[7] and co-detection by indexing (CODEX),[8] often suffer from hurdles like staining inconsistency, substantial time and cost requirements,[9,10] as well as the

*Corresponding author.
**Corresponding author.
***Corresponding author.
E-mail addresses: zixia@stanford.edu (Z. Zhou), yumjiang@wakehealth.edu (Y. Jiang), lei@stanford.edu (L. Xing).
[e]Contributed equally as first authors.

Research in context

Evidence before this study

Before undertaking this study, we conducted a comprehensive review of the existing literature and available evidence on standard multiplexed immunofluorescence (mIF) staining, virtual mIF staining, and related deep learning techniques in pathological imaging. We searched multiple databases, including PubMed, HTAN, and TCIA, focusing on studies published up to April 2023. The search terms included combinations of "virtual staining", "multiplexed immunofluorescence", "deep learning", "autofluorescence (AF)", "DAPI imaging", and "gastric cancer prognosis".

We included studies that explored the use of deep learning for virtual staining and image-to-image translation in pathology. The evidence reviewed included works demonstrating the potential of generative adversarial networks (GANs) for cross-modality super-resolution in fluorescence microscopy, as seen in Wang et al. (2020), and the application of deep learning to bright-field holography by Wu et al. (2021). Additionally, research by Zhang et al. (2022) on multiplexed H&E, Jones silver, and Masson's trichrome staining using AF images was considered, as well as studies by Sun et al. (2021) and He et al. (2022) on *in silico* immunohistochemical staining from H&E images. The quality of these studies varied, with some presenting robust experimental designs and extensive computational validation, while others were more preliminary in nature. Notably, the clinical reliability of virtual staining was often limited, and the usability of these methods in routine pathological workflows was not consistently validated. While some studies demonstrated the potential for deep learning models to produce high-quality virtual staining images, the translation of these methods into clinically usable pathological modalities remained a significant challenge.

The pooled estimate from the meta-analysis of these studies suggested that deep learning models show promise in generating high-quality virtual staining images comparable to traditional methods. However, challenges remained in achieving consistent performance across diverse biomarkers and tissue types.

Added value of this study

This study introduces the Multimodal-Attention-based virtual mIF Staining (MAS) system, which significantly advances the field by employing a deep learning model to generate high-quality virtual mIF images from dual-modal non-antibody-stained fluorescence imaging, specifically AF and DAPI imaging. The MAS system utilises self- and multi-attention mechanisms to accurately predict multiple survival-associated biomarkers in gastric cancer, providing a cost-effective and rapid alternative to traditional mIF techniques.

Our findings add substantial value to the existing evidence by demonstrating that the MAS system can achieve prognostic accuracy comparable to standard mIF staining. We validated the system using 180 pathological slides from 94 gastric cancer patients, showing consistent performance across both cancerous and non-cancerous gastric tissues. The inclusion of seven key gastric cancer biomarkers (CD3, CD20, FOXP3, PD1, CD8, CD163, and PD-L1) in the study highlights the system's versatility and potential clinical applicability.

Implications of all the available evidence

The integration of our MAS system into clinical practice could revolutionise the workflow for multiplexed staining, significantly reducing both time and costs. This advancement has profound implications for personalised medicine, enabling more efficient and accessible prognostic evaluations for gastric cancer patients and potentially for other cancers as well.

The ability to rapidly generate reliable mIF images from easily obtainable AF and DAPI slides can facilitate broader adoption of multiplexed staining in clinical and research settings. This method alleviates the high costs and labour-intensive nature of traditional mIF techniques, promoting more widespread use in routine diagnostics and large-scale studies.

Looking ahead, we aim to design improved networks that can incorporate precise and stable nuclear information from DAPI through pre-training. This advancement will enable our system to generate real-time mIF images directly from AF inputs during the testing phase, without any prior staining. Future research should also focus on expanding the application of the MAS system to other biomarkers and cancer types, further validating its generalisability and robustness. Additionally, exploring the integration of different wavelengths in AF imaging could enhance the system's performance and accuracy. Continued development and optimisation of deep learning models for virtual staining will undoubtedly contribute to advancing precision medicine and improving patient outcomes.

fading of stained slides over time (Fig. 1a). Given these limitations, there is an urgent demand for a rapid and cost-effective solution to overcome these limitations and facilitate its clinical application.

Recent advancements in generative artificial intelligence (AI) have shown significant potential in image-to-image translation,[11–15] particularly in pathological image generation.[16–19] Several studies have focused on improving the resolution and quality of pathological imaging across different modalities. For instance, Wang et al.[20] employed a generative adversarial network (GAN) to achieve cross-modality super-resolution in fluorescence microscopy, enhancing low-resolution images to match high-resolution ones and democratising access to advanced imaging techniques. Similarly, Wu et al.[21] applied cross-modality deep learning to bright-field holography, using a GAN to achieve artifact-free bright-field contrast in holographic images. Other researchers

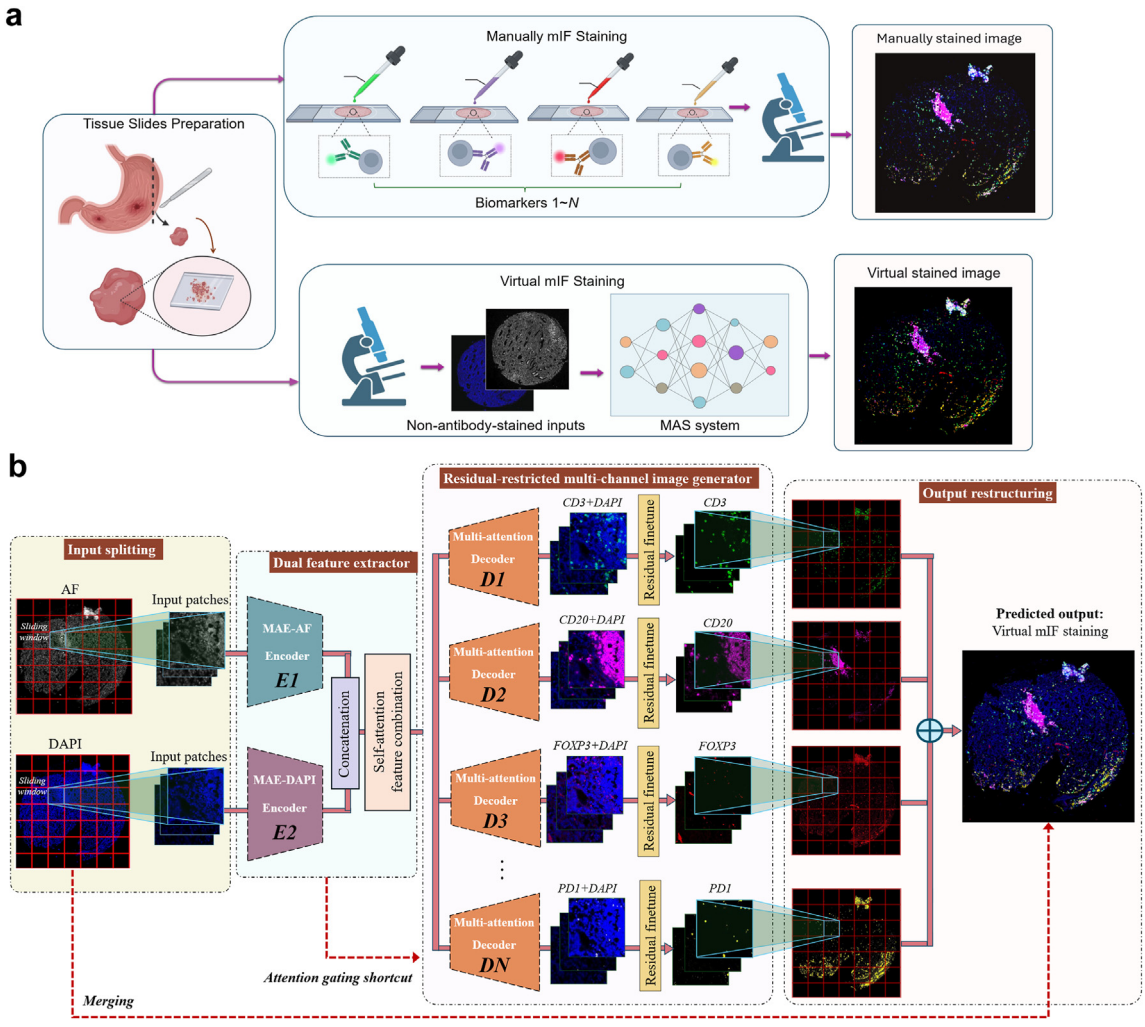


Fig. 1: Overview of the proposed MAS system. (**a**) Pipeline comparison between the standard manually mIF staining and the virtual mIF staining by MAS system. (**b**) Architecture of MAS system.

have investigated regression-based models to learn the relationships between different pathological imaging patterns or modalities for various applications.[22] For example, Zhang et al.[23] developed a method for multiplexed H&E, Jones silver, and Masson's trichrome staining using a single network that takes autofluorescence (AF) images and a digital staining matrix as inputs. Jo et al.[24] introduced label-free multiplexed microtomography, using deep learning to decode refractive index measurements into 3D fluorescence tomograms. Further studies have explored deep learning techniques for antibody-related virtual staining. Among them, Sun et al.,[25] He et al.,[26] and Zhang et al.[27] Utilised deep learning to produce *in silico* immunohistochemical (IHC) staining from haematoxylin and eosin (H&E) stained tissue images. Additionally, research has explored the potential of deep learning to digitally generate specific antibody stains from AF imaging, such as virtual IHC HER2 staining using AF images.[28] Efforts have also been made to directly predict multiplexed fluorescent labels in unlabelled transmitted-light z-stacks.[29] In addition, to enhance the accessibility and cost-effectiveness of high-plex mIF insights, Wu et al. developed a machine learning framework capable of generating *in silico* 40-plex CODEX data from a standard 7-plex mIF panel.[30] These studies underscore the presence of inherent and unexplored information related to biomarkers in non-antibody-stained fluorescence images of tissues and cells.

Aiming to devise a rapid, economical, and reliable mIF framework for gastric cancer prognosis, we introduced a multimodal-attention-based virtual mIF staining (MAS) system, developed to efficiently and reliably produce virtual mIF staining from non-antibody-stained fluorescence imaging. The MAS model is built upon an end-to-end generative convolutional neural network

(CNN) and utilises AF and 4′,6-diamidino-2-phenylindole (DAPI) slides as inputs to generate mIF staining for gastric cancer related biomarkers (Fig. 1b). The model's feature extractors, fortified with pretrained masked auto-encoders (MAEs) (Fig. 2a) and a self-attention combination strategy (Fig. 2b), adeptly extract antigen-label-related features and precisely locate targeted cells. Employing a composite loss function minimisation, parallel multi-attention decoders concurrently predict a set of images for varied biomarkers (Fig. 2c). The MAS system's efficacy was assessed using a cohort of 94 patients with gastric cancer, successfully achieving virtual staining for seven gastric cancer survival-associated biomarkers (CD3, CD20, FOXP3, PD1, CD8, CD163, and PD- L1) in cancerous and non-cancerous gastric tissues, with quality equivalent to manually stained reference images. Furthermore, the prognostic prowess of the virtual mIF images was validated, matching results derived from manually stained mIF images. The MAS system stands out by enabling the rapid, cost-effective generation of reliable mIF staining. Through its consistent and efficient performance, it holds the potential to redefine mIF staining paradigms, offering a pathway towards personalised medicine without imposing substantial time and financial burden.

Methods

Ethics

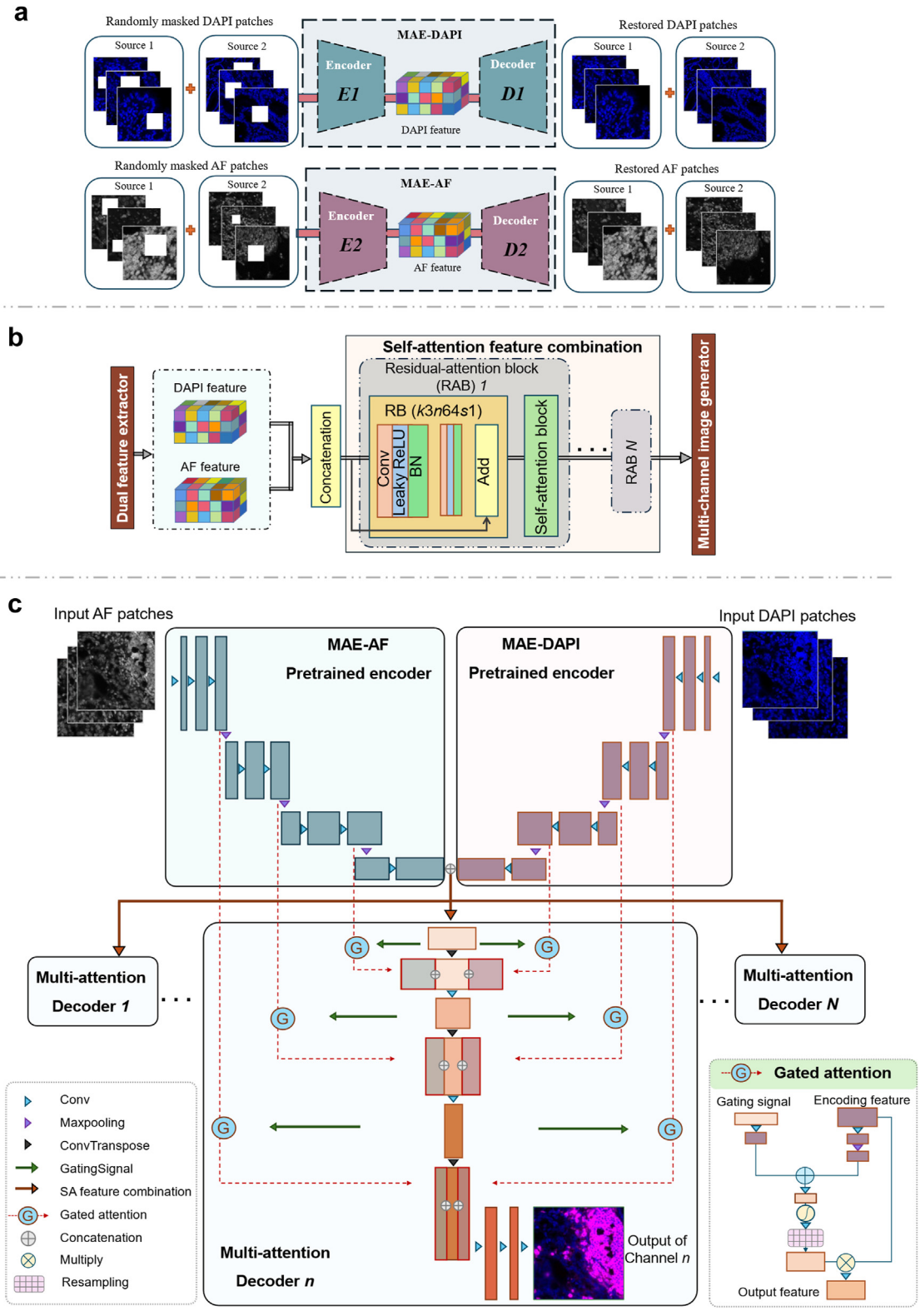
This study was approved by the institutional review board of Nanfang hospital (NFEC2017171). Original data was collected from Nanfang Hospital (Guangdong, China) from July to December 2010, and the survival information was followed up and recorded until December 2016. All manual staining processes were conducted at the Shanghai Outdo Biotech Company (Shanghai, China). Considering the retrospective nature of this study, the requirement for informed consent was waived.

Data preparation

The detailed data preparation procedure, including the tissue microarray and training data staining rules, is provided in the Supplementary Material, Table S4, and Table S5. Specifically, for the AF images: excitation wavelength = 360 nm; emission wavelength = 420 nm. For the DAPI images: excitation wavelength = 350 nm; emission wavelength = 450 nm. These wavelengths were selected to optimise the visualisation of specific tissue components and biomarkers. The chosen wavelengths for AF allow for the effective imaging of endogenous fluorophores such as collagen and elastin, which are commonly present in gastric tissue. After data collection, we obtained experimental data from 94 patients with gastric cancer. During the data collection, sex was self-reported by study participants. Since our research focuses on the efficacy of the staining network and its clinical reliability in gastric cancer prognosis, sex was considered as a covariate of non-interest. Pathological slides of both cancer and noncancer tissues are selected. To acquire aligned training pairs, we first generated DAPI and AF images of each tissue sample and then simultaneously stained them with CD3, CD20, FOXP3, and PD1 labels. Finally, we obtained 180 aligned slide pairs, each possessing dimensions of 3616 × 3612. The same procedure was executed to obtain another set of 180 aligned slide pairs for the CD8, CD163, PD-L1 stained slides, along with their corresponding DAPI and AF slides. In network training, we used data from 80 patients as the training set and data from the remaining 14 patients as the independent testing set for the original task dataset. For the transfer learning task, we employed an additional dataset from the same patients but with different scanning and staining times. This dataset followed the same train-test split strategy as the original task. The independent testing set ensures the integrity of our evaluation process. We implemented a sliding window strategy to generate patches from the whole slides. For the training set involving predictions of CD3, CD20, FOXP3, and PD1 biomarkers, the sliding window yielded 26,563 patches, whereas for the training set of CD8, CD163, and PD-L1 predictions, we obtained 10,240 patches in total. All patches were resized to 256 × 256, and normalised to a scale of 0–1. During network training for both the original task and the transfer learning task, 20% of the original training set was randomly selected as the validation set, while the remaining 80% was used as the actual training set. A standard data augmentation process[31] was conducted to increase the diversity of the training data and enhance the robustness of network training. By applying transformations such as shifting, rotating, shearing, zooming, and flipping with random parameters during each training epoch, our augmentation strategy effectively broadens the training landscape. This approach helps prevent the model from overfitting to the peculiarities of the training set.

Experimental setup

During the training process, the two MAE models were pretrained using the stochastic gradient descent (SGD) optimiser. The initial learning rate was 10^{-4} with a momentum of 0.9 and a clip value of 5.0. Convergence of the MAE models was considered achieved when the loss value on the validation set did not decrease for 20 consecutive epochs. For training the MAS network, we employed the Adam optimiser with an initial learning rate of 10^{-4} and a weight decay of 10^{-2}. Similarly to the MAE models, we determined the convergence status of the MAS by monitoring the loss value on the validation set. The experiments were performed in Linux operating system (Ubuntu 16.04 LTS), with Python 3.7. The training process was conducted using Keras with Tensorflow as the backend. In addition, an NVIDIA Tesla V100 DGXS graphic processing unit (GPU) was adopted to increase the training speed.


a
Randomly masked DAPI patches
Source 1
Source 2
MAE-DAPI
Encoder
E1
Decoder
D1
DAPI feature
Restored DAPI patches
Source 1
Source 2
Randomly masked AF patches
Source 1
Source 2
MAE-AF
Encoder
E2
Decoder
D2
AF feature
Restored AF patches
Source 1
Source 2
b
Dual feature extractor
DAPI feature
AF feature
Concatenation
Self-attention feature combination
Residual-attention block
(RAB) 1
RB (k3n64s1)
Conv
Leaky ReLU
BN
Add
Self-attention block
RAB N
Multi-channel image generator
c
Input AF patches
Input DAPI patches
MAE-AF
Pretrained encoder
MAE-DAPI
Pretrained encoder
Multi-attention
Decoder 1
Multi-attention
Decoder N
Multi-attention
Decoder n
Output of
Channel n
Conv
Maxpooling
ConvTranspose
GatingSignal
SA feature combination
Gated attention
Concatenation
Multiply
Resampling
Gated attention
Gating signal
Encoding feature
Output feature

MAS system details

a) Deep learning architecture establishment

The network within the MAS system melds four submodules: the input splitting module, dual feature extractor module, multi-channel image generator module, and output restructuring module (Fig. 2).

The input splitting module utilises a sliding window of size 256 × 256 to traversing over the slides, extracting autofluorescent and DAPI patches from the corresponding whole slides. Concurrently, the dual feature extractor employs AF-based and DAPI-based MAEs to capture relevant features from the inputs. Each MAE is pretrained using AF images and DAPI images, effectively harnessing targeted characteristics from both input modalities (Fig. 2a). The MAEs undergo training with a blend of data from both panel 1 and panel 2, incorporating a random masked mechanism, wherein each patch receives a random-sized, randomly-located mask during MAE training in each epoch, with a new mask generated in subsequent epochs. The encoder outputs from the MAEs intertwine through a concatenation layer, illustrated in Fig. 2b.

Subsequently, multiple residual blocks and attention blocks are employed to integrate the multimodality features, incorporating both cytoplasm and nucleus information. The multimodality features extracted from the two MAEs, named MAE-AF and MAE-DAPI, are then concatenated and combined by a self-attention feature combination block inspired by Zhang et al.'s research.[32] The self-attention block, composed of several residual-attention sub-blocks, allows the network to spotlight essential parts of the double-expanded feature dimension.

Moreover, drawing upon the proven efficacy of self-attention in image-to-image generation and the robust feature extraction capability of the residual attention module, attention blocks are strategically incorporated (see Figure S13). Each attention block consists of a trunk branch and a mask branch. The mask branch, comprising max pooling layers, residual blocks, and upsampling layers, generates an attention mask that highlights relevant features and suppresses noise. The attention system's dexterity in directing the network towards the input's most informative aspects has proven to be instrumental, offering a nuanced understanding and visualisation of the complex, multimodal biological data.

Subsequently, the multi-channel image generator branches out into N image generating channels, facilitating parallel prediction of pathological slides, each stained with N distinct biomarkers. In this study, we set N to four, generating staining images for the CD3, CD20, FOXP3, and PD1 biomarkers. The output from the dual feature extractor is then fed into four parallel multi-attention decoders, each following the structure showcased in Fig. 2c. The multi-attention decoder replaces conventional skip connection with a gated attention feature connection.[33] This mechanism utilises the decoder feature as a gating signal to selectively prune low-level encoder features from MAE-AF and MAE-DAPI, assuring that insightful information is effectively extracted from the multimodality inputs. The gating coefficient is obtained using additive attention, which is defined as follows:

$$q_{att}^{l} = \psi^{T}\left(\sigma_1\left(W_x^{T} x_i^{l} + W_g^{T} g_i + b_g\right)\right) + b_\psi \tag{1}$$

$$\alpha_i^{l} = \sigma_2\left(q_{att}^{l}\left(x_i^{l}, g_i; \Theta_{att}\right)\right) \tag{2}$$

where $\sigma_2(x_{i,c}) = 1/1+\exp(-x_{i,c})$. Θ_{att} denotes the entire parameters in the gated attention calculator. To be specific, Θ_{att} includes linear transformations W_x^T, W_g^T, ψ and bias terms b_ψ, b_g.

The multi-attention decoders produce positive signals stained with different labels, utilising DAPI images as the background. This design choice is motivated by the precise positional information of cell nuclei provided by DAPI, which allows the network to accurately determine the location of each cell and contribute to precise staining positioning. Afterwards, outputs from the multi-attention decoders are directed to the residual finetune block, which deducts the input DAPI image from the decoders' output slides, deriving the residual signal. This signal is then refined via three convolutional layers to formulate the final staining prediction. This system allows for the parallel generation of desired predictions, accommodating flexibility in producing multiple staining outcomes.

b) System training and hyperparameter optimisation

The training process of MAS system contains two steps. The first step is the pre-training of MAE-AF and MAE-DAPI. In this stage, each MAE uses MSE as the loss function and applies randomly sized and randomly-located masks to the training patches in each epoch. The objective of the network is to generate the unmasked patches that correspond to the input patches after going through the encoder-decoder structure with multiple convolutional layers. The composite loss function can be defined as follows:

$$l_{MAE} = \frac{1}{M}\sum_{i=1}^{M}\left\| I_i^{ref} - \mathrm{MAE}\left(I_i^{in}\right)\right\|_2^2 \tag{3}$$

Fig. 2: MAS system method details. **a**, Detailed architecture of masked auto-encoders, where the encoders are utilised as powerful feature extractors in the MAS system. **b**, Self-attention feature combination structure. c, Gated attention details in multi-attention decoder, which presents the detailed gated attention connection mechanism used in the MAS system.

where MAE(I_i^{in}) and I_i^{ref} denote the predicted virtual mIF slides of MAS in the training batch and its corresponding standard manually stained slides of the *i*-th sample, respectively, and *M* is the training batch size. Compared to conventional auto-encoder feature extractors, the MAE compels the network to predict unknown masked features using their surrounding information. As a result, the MAE demonstrates enhanced generalisation capabilities and possesses superior feature capturing and contextual feature prediction abilities.

Next, we begin by extracting the pretrained weights from the encoders of MAE-AF and MAE-DAPI and then loading them into the corresponding layers of the MAS model. The remaining parts of the model are initialised using Glorot Uniform initialisation, which generates initial values from a Uniform distribution. The loss function of the MAS network consists of two components: constraints on the outputs of the multi-attention decoder module and the residual finetune module. Our objective is to minimise the MSE between the predicted virtual mIF signals and the corresponding manually stained signals, aiming to reduce the differences between them. The final loss can be formulated as follows:

$$l_{MAS} = \sum_{n=1}^{N} l_{MAS}^{n} \tag{4}$$

$$l_{MAS}^{n} = \frac{1}{M}\left[\sum_{i=1}^{M}\left\| I_i^{r1} - MA\left(I_i^{AU}, I_i^{DA}\right)\right\|_2^2 + \sum_{i=1}^{M}\left\| I_i^{r2} - RF\left(MA\left(I_i^{AU}, I_i^{DA}\right)\right)\right\|_2^2\right] \tag{5}$$

where *M* is the training batch size, *N* is the number of mIF labels the MAS predicted, l_{MAS}^{n} denotes the sub-loss for *n*-th staining prediction. $MA(\cdot)$ and $RF(\cdot)$ represent the calculations of multi-attention decoder and residual finetune modules, respectively. I_i^{AU}, I_i^{DA}, I_i^{r1} and I_i^{r2} represent the AF input, DAPI input, referenced mIF manual-stained image covered on DAPI image and the pure referenced mIF stained image, respectively.

c) Implementation of transfer learning for unseen label adaption

To ensure the effective adaptation of the model in the proposed MAS system to new immunofluorescent labels, we employed transfer learning, a widely used technique in deep learning. This process involved two stages. In the first stage, we transferred the pretrained weights of the dual feature extractor, which were already trained on the original task, to the new label training. We then froze these weights and trained the parameters of the residual-restricted multi-channel image generator from scratch until the network reached convergence.

In the second stage, we unfroze the dual feature extractor and finetuned the entire MAS system with a small learning rate until convergence was reached again. This approach allowed the network to adjust its parameters and learn more specialised features for the new labels. As a result, the MAS system could rapidly adapt to unseen immunofluorescent labels with less additional training data and faster convergence speed. Furthermore, this approach reduced the likelihood of overfitting the network to a specific set of labels.

Model complexity

The MAS model, which simultaneously predicts CD3, CD20, FOXP3, and PD1, has 9,021,688 parameters and 4.5219×10^{11} FLOPs. The MAS model's training time is approximately 72 h. During the testing stage, the processing time to stain a whole-slide image is around 3.3 s with a GPU and about 38 s with a CPU. Similarly, staining for CD8, CD163, and PD-L1 takes around 2.6 s with a GPU and approximately 30 s with a CPU.

Quantification and statistical analysis

To evaluate the performance of the different methods, we employed various assessment techniques, including similarity-based metrics, patch-based clinical observation assessment, and survival prediction analysis.

a) Computational-similarity-based evaluation

The similarity-based metrics used in the evaluation included the PSNR and the SSIM. These metrics provide measures of the pixel-based and perceptual-based similarities between the predicted virtual mIF slides and the standard manually stained images (without DAPI as background), respectively. The PSNR is calculated using the following formula:

$$\mathrm{PSNR} = 10 \times \log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right) \tag{6}$$

where MSE represents the mean squared error calculated between the virtual mIF slides of the i-th input in the testing dataset and its corresponding standard manually stained slides. The SSIM is calculated by.

$$\mathrm{SSIM} = \frac{(2\mu_M\mu_A + C_1)(2\sigma_{A,M} + C_2)}{(\mu_M^2 + \mu_A^2 + C_1)(\sigma_M^2 + \sigma_A^2 + C_2)} \tag{7}$$

where μ_A, μ_M, σ_A, σ_M and $\sigma_{A,M}$ denote the means, standard deviations and covariance of the virtual mIF slides and the standard manually stained slides, respectively. $C1$ and $C2$ are defined values used to stabilise the calculation.

b) Patch-based pathologists' observation assessment

To evaluate the density of stained immune cells, the nucleated stained cells in each core were quantified and expressed as the number of cells per core. This evaluation was performed by 2 gastroenterology pathologists

who were blinded to the clinical outcome. In cases where differences arose between the two primary pathologists, a third pathologist was consulted to reach a consensus. Subsequently, we evaluated the observed results using Pearson and Spearman correlation analyses to measure the linear and monotonic relationships, respectively, between the virtual mIF patches and the manually stained patches. By employing both correlation measures, we ensured that our findings are robust and not solely dependent on the assumption of linearity.

c) Survival prediction analysis

Among the 94 patients with gastric cancer, 86 have both gastric cancer tissues and matched noncancer samples, while 8 patients only have gastric cancer samples. For the survival analysis of patients with gastric cancer, we only used the tumour samples, as all patients have these tumour samples. We assessed the association of each biomarker's expression with overall survival (OS). OS was defined as the time from surgery to death from any cause. Survival curves were generated according to the Kaplan–Meier method and compared using the log-rank test. Statistical analysis was conducted with R software (version 4.1.0). A two-sided p value < 0.05 was considered statistically significant. The cutoff score for the density of each biomarker was the medium number of stained immune cells. The combined positive score (CPS) was defined as the total number of PD-L1 positive cells (tumour, lymphocytes, and macrophages) divided by the total number of viable tumour cells multiplied by 100. CPS was categorised as high (CPS $\geq$10), intermediate (10 > CPS $\geq$1), and low (CPS <1). We used the multivariable Cox regression model to select the biomarkers, and then constructed an integrated score.

In addition, we performed the subgroup analysis stratified by clinicopathological factors. We also conduct multivariate Cox analyses using the clinical factors as covariates, including sex, age, tumour grade, tumour size, tumour location, depth of invasion, and lymph node metastasis, distant metastasis, TNM stage.

Statistics

In this study, various statistical methods were employed to ensure the robustness and validity of the findings. Sample size determination was conducted using data from 94 patients with gastric cancer. Of these, 80 patients were considered as training set, and 14 patients formed an independent testing set. Randomisation was implemented by randomly selecting 20% of the original training set as the validation set for network training. Additionally, during patch-based pathologists' observation assessment, two gastroenterology pathologists, blinded to clinical outcomes, evaluated the density of stained immune cells. Pearson and Spearman correlation analyses were conducted to measure the linear and monotonic relationships between virtual and manually stained patches. For survival prediction analysis, Kaplan–Meier analysis and log-rank tests were utilised to demonstrate the prognostic value of the biomarkers. The cutoff score for each biomarker was set as the median number of stained immune cells. Multivariable Cox regression models were then used to select biomarkers and construct an integrated score, accounting for various clinical factors.

Role of funders

The funders had no role in the study design, data collection, data analyses, interpretation or writing of the report; the findings and conclusions contained herein are those of the authors and do not necessarily reflect positions or policies of the aforementioned funding bodies.

Results

The proposed MAS system enables high-performance virtual mIF staining

For training of the MAS system, we collected 180 slides of tissue samples containing both gastric cancer (94 samples) and noncancer (86 samples) from 94 patients with gastric cancer. During the data preparation phase, we obtained AF and DAPI images for each tissue sample, and the mIF images with four biomarkers (CD3, CD20, FOXP3, and PD1) from the same sample. Here, the four immune cell biomarkers are commonly employed in evaluating patients' prognosis and treatment outcomes.[34]

We illustrate the predicted staining slides of cancer and noncancer tissue cases in Fig. 3a and b. The virtual mIF images exhibit remarkable similarity to the corresponding manually stained images, demonstrating the capability of the proposed MAS system in achieving automated staining of both cancer and noncancer tissues.

Patch-based observation assessment of pathologists was also conducted, where 100 sets of patches were randomly selected for each biomarker. Pathologists evaluated and recorded the number of true positive cells in both the predicted virtual mIF stained patches and the manual-stained referenced mIF patches. Scatter plots were generated to visualise and compare the results (Fig. 3c). The scatter plots clearly depict a distinct thin diagonal line, indicating a strong positive correlation between the predicted virtual mIF patches and the standard manually stained patches. The Pearson correlation coefficient (PCC) values for the CD3, CD20, FOXP3, and PD1 labels are 0.884, 0.970, 0.981, and 0.963, respectively. The corresponding p-values in t-test for the CD3, CD20, FOXP3, and PD1 labels are 2.37×10^{-41}, 2.89×10^{-64}, 1.11×10^{-93}, and 1.30×10^{-75}, respectively. The Spearman correlation coefficient values for the CD3, CD20, FOXP3, and PD1 labels are 0.809, 0.959, 0.943, and 0.879, respectively. The

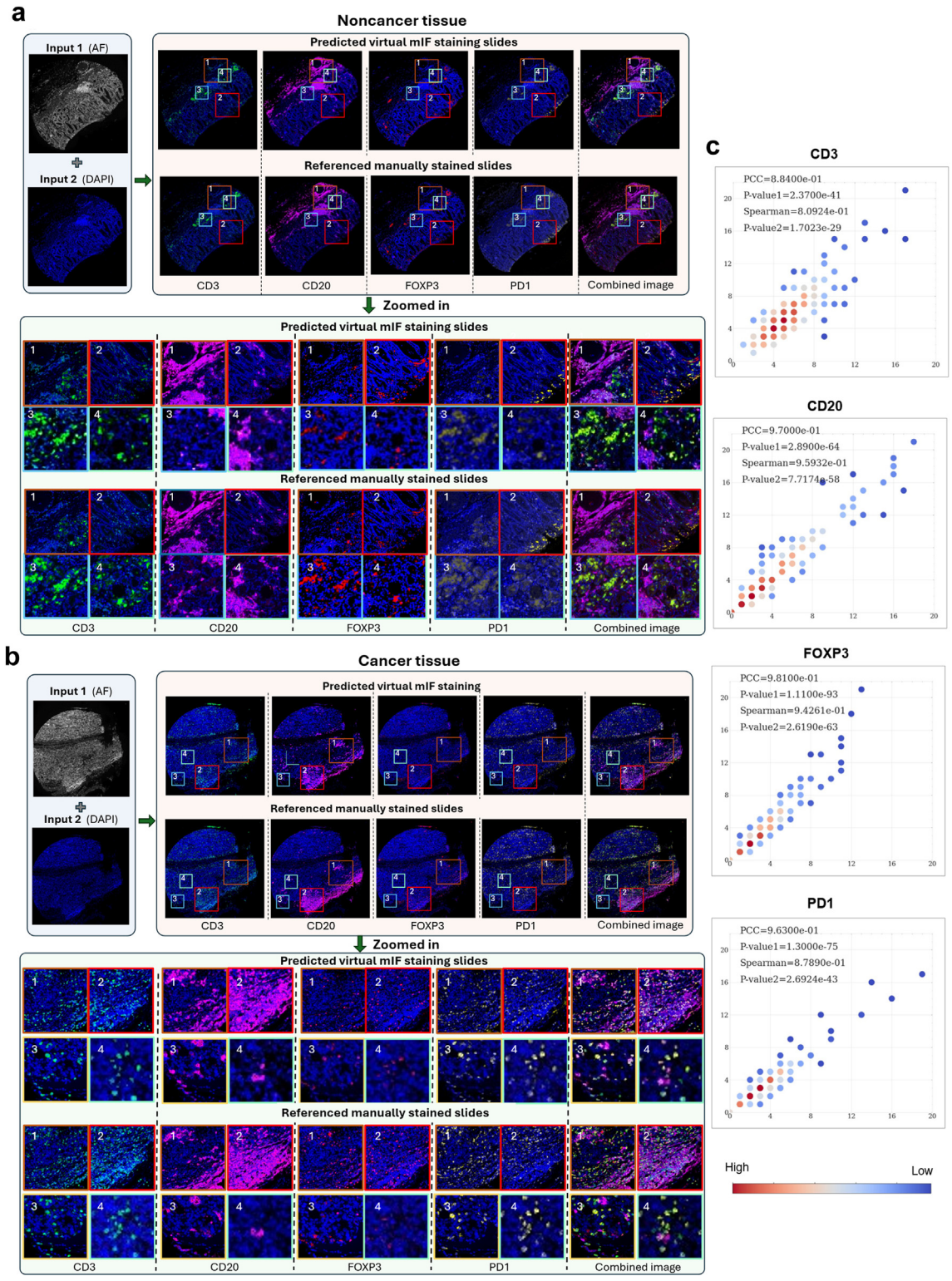


Fig. 3: Illustrative examples of the predicted staining slides for noncancer cases (a) and cancer cases (b). For each case, we present three elements: the inputs to the MA model, denoted as non-antibody-stained fluorescence images; the outputs generated by the MA model, being the virtual mIF stained images; and the standard manually mIF stained images. The results independently display CD3, CD20, FOXP3, PD1, along

corresponding p-values in approximate t-test for the CD3, CD20, FOXP3, and PD1 labels are 1.70×10^{-29}, 7.72×10^{-58}, 2.62×10^{-63}, and 2.69×10^{-43}, respectively. These results provide compelling evidence that the virtual mIF images obtained from the MAS system exhibit a high degree of similarity to the standard manually stained slides when examining individual cells.

Quick adaptation of MAS model for virtual staining of other biomarkers via transfer learning

The MAS system is generalisable to unseen labels through a transfer learning step (Fig. 4a). To ensure effective adaptation of the MAS system to new labels, the transfer learning process was performed in two phases. First, the pretrained weights of the dual feature extractor, obtained from the original task, are transferred and frozen for the new label prediction task. The remaining parameters in the residual-restricted multi-channel image generator are trained from scratch until convergence is achieved. Second, the dual feature extractor is unfrozen, and the MAS system is finetuned with a lower learning rate until it converges again. In this way, the MAS system can rapidly adapt to any unseen mIF staining task.

The transfer learning dataset included DAPI and AF slides, as well as stained slides with CD8, CD163, and PD-L1 labels. The results of transfer learning are presented in Fig. 4b, showcasing the capability of the MAS system in mIF staining for external biomarkers. Among the three new labels, the proposed approach demonstrates the best performance in predicting the PD-L1 label, potentially attributed to the strong correlation between PD-L1 and PD1, which is one of the training labels in the original network. Concerning CD163, the MAS system accurately predicts the location of positive signals, although the presence of diffuse false positive signals in the referenced images of the training data slightly affects the network training, resulting in reduced sharpness of the output image. In the case of CD8 staining, the abundance of numerous small target positive signals poses a challenge for prediction, leading to some loss of detail in the transferred results. Overall, the transfer learning strategy employed by the MAS system generates staining slides that exhibit high similarity to the referenced images, thus confirming its robust generalisation capability.

Patch-based observation assessment by pathologists was also conducted to evaluate the transferred results. The scatter plots comparing the predicted virtual mIF patches with the manually stained patches are shown in Fig. 4c. The PCC values for the CD8, CD163, and PD-L1 labels are 0.930, 0.944, and 0.888, respectively. The corresponding p-values in t-test for these labels are 9.05×10^{-58}, 2.62×10^{-49}, and 1.45×10^{-35}, respectively. The Spearman values in approximate t-test for the CD8, CD163, and PD-L1 labels are 0.931, 0.934, and 0.923, respectively. The corresponding p-values in approximate t-test for these labels are 2.07×10^{-58}, 6.49×10^{-46}, and 2.71×10^{-43}, respectively. The results of patch-based observation assessment (Fig. 5a) provide strong evidence that the transferred learning results of the MAS system are highly correlative to the manually stained images, particularly when examining individual cells.

Survival prediction of patients with gastric cancer by using the virtual mIF images

The MAS system was used to perform end-to-end predictions of the gastric cancer related biomarkers at patient level. We found the virtual mIF images generated by the MAS system not only exhibit high resemblance to the manually stained images, but they also offer reliable information for downstream applications such as survival prediction. The association between the selected biomarkers (CD3, CD20, FOXP3, PD1, CD8, CD163, and PD-L1) and prognosis is well established and has been validated in previously published studies.[35,36] We utilise this downstream task to evaluate the clinical consistency of pathologists' prognostic predictions based on virtual stained images compared to manually stained reference images. During the prognostic analysis, we first classified each patient as high or low expression group based on the numbers of positive cells. Then, we performed Kaplan–Meier analysis[37] with log-rank test to show the prognostic value of the generated biomarkers (Fig. 5b and c). Our results revealed that patients in the CD3-high group exhibited significantly better OS compared to those in the CD3-low group ($p < 0.001$). This is consistent with the analysis results observed from the standard manually stained images. We expanded our analysis to other generated stained biomarkers, namely CD20, FOXP3, PD1, CD8, CD163, and PD-L1. Among these biomarkers, CD20, CD8, PD1, and CPS score demonstrated a significant association with longer OS, while CD163 and FOXP3 showed a tendency towards shorter OS. In a multivariable Cox proportional hazards regression analysis, CD3, FOXP3, CD163, and CPS score were identified as robust prognostic biomarkers, with corresponding regression

with the consolidated multiplexed outcomes. To clearly depict the intricate details, we enlarged four distinct sections, located below the comprehensive slide images. (c) Scatter plots of the virtual mIF patches and manually stained patches for CD3, CD20, FOXP3, and PD1 labels. The colour gradient shows the density of the dots. The X-axis and Y-axis denote the count of true positively stained cells on the virtual mIF patches and the manually stained mIF patches, respectively. The colour gradient shows the density of the dots.

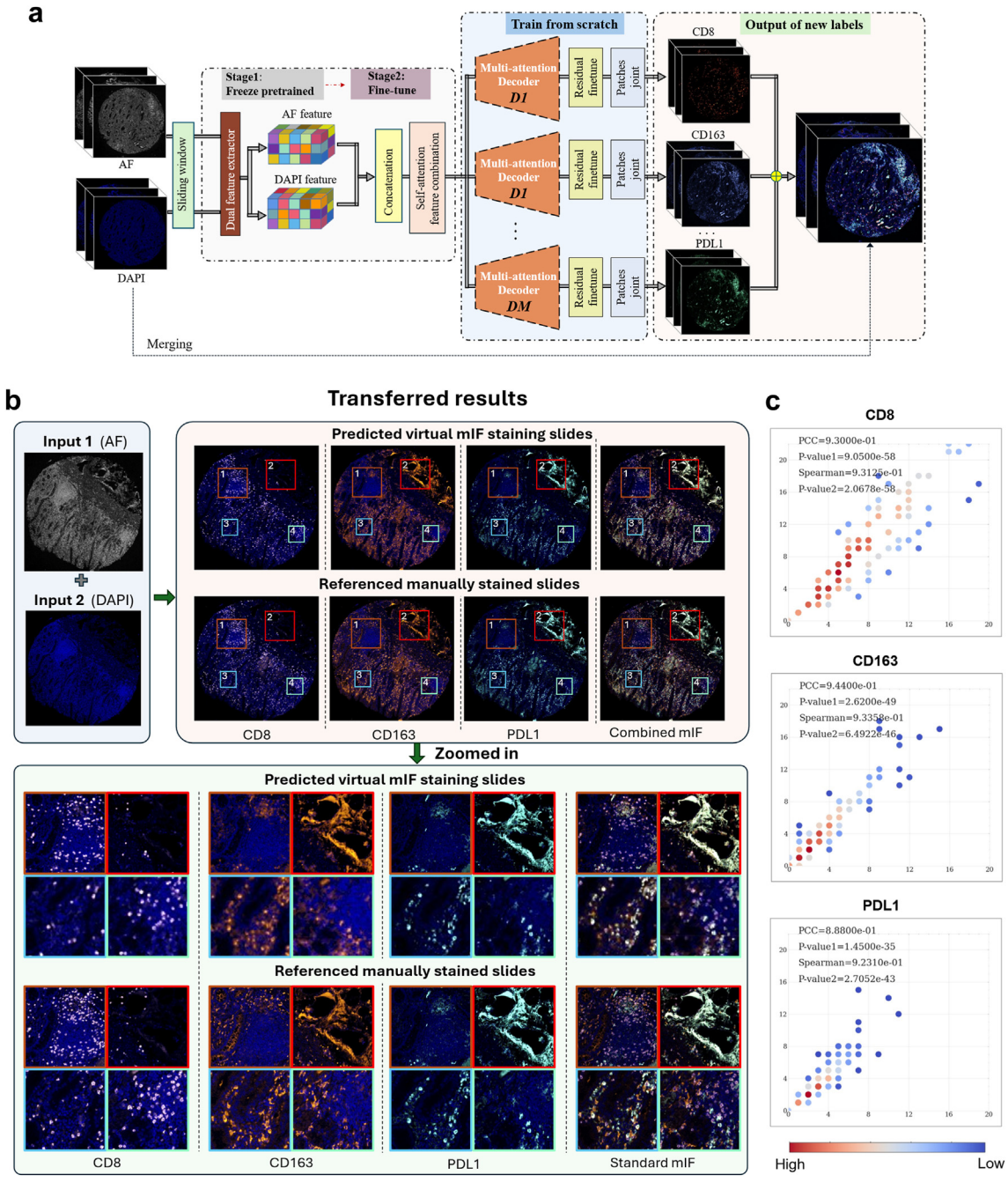


Fig. 4: Illustration and results of transfer learning strategy in the MAS. (a) Details of the transfer learning strategy. To facilitate quick adaptation of a trained MAS system to unseen new biomarkers, we implement a two-phase transfer learning process. The framework's generalisability is highlighted by its flexible output channel. (b) and (c) Illustrative examples and scatter plots results of the transferred results.

coefficients of 1.299778, –0.911491, –0.804614, and 0.49202, respectively. Utilising these coefficients, we calculated an Integrated Score for each patient using the formula: Integrated Score = 1.299778 × CD3–0.911491 × FOXP3–0.804614 × CD163 + 0.49202 × CPS. The Integrated Score exhibited a significant association with longer OS ($p < 0.0001$). These findings provide compelling evidence that the generated stained slides can yield clinical prognostic results comparable to those obtained from manually stained images. In the subgroup analysis stratified by clinicopathological factors, both the integrated scores of virtual staining and

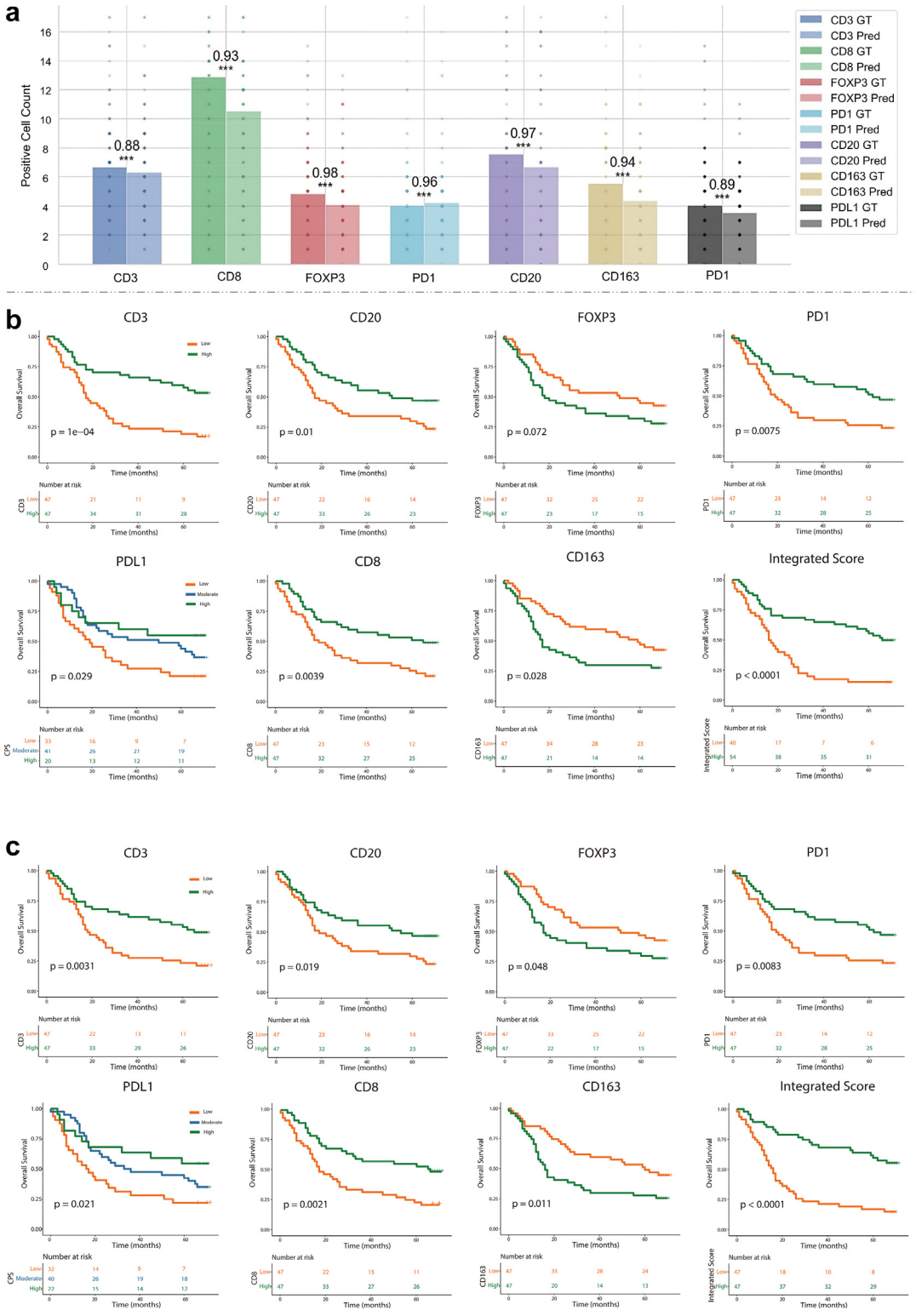


Fig. 5: (a) Overall performance through the observation assessment. To compare the similarity between the observation results between the manually stained images and the virtual stained images, we calculate the Pearson correlation value and the p-value (using a t-test) for CD3, CD20, FOXP3, PD1, CD8, CD163, PD-L1, noted that * represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$. (b) Kaplan–Meier curves according to predicted virtual stained slides labelled by different biomarkers for survival analysis of patients with gastric cancer. (c) Kaplan–Meier curves according to manually stained slides labelled by different biomarkers for gastric cancer survival. Each plot in (b) and (c) shows the probability that a subject will survive up to time *t*.

manually staining were still associated with patients' OS. Moreover, multivariate Cox regression analysis after adjustment for clinicopathological variables revealed that both the integrated scores of virtual staining and actual staining remained powerful prognostic factor (see Figure S11, Figure S12, Table S6, and Table S7).

Discussion

Generative AI has emerged as a promising strategy for image reconstruction, generation, and translation.[12,38] In general, existing research in pathological image generation can be divided into two categories. The first centres around generating high-quality functional stained images from low-quality functional stained images.[19] The second aims to generate functional stained images from non-functional stained images.[26,27,29] In both categories, however, significant efforts are demanded to prepare the input images, raising concerns about the practicality of these methods for widespread clinical applications. In this work, we have presented a generative MAS deep learning strategy to generate multiplexed mIF images from low-cost and easily attainable AF and DAPI images. The quality of the mIF images and downstream prognosis are demonstrated by using 180 slides of pathological images from 94 patients with gastric cancer. By harnessing advanced deep learning techniques and integrating modules such as the input splitting module, dual feature extractor module, multi-channel image generator module, and output restructuring module, the proposed MAS system empowers rapid mIF staining from unlabelled slides. It is intriguing that the technique enables nearly instant multiplexed staining, resulting in significant cost reduction and improved workflow.

A special case analysis was conducted as shown in Fig. 6a. In *Case 1*, the whole slides contain a large number of densely packed, relatively small positive cells for all four biomarkers. Our MAS method generally locates the positions of positive cells, but for CD3 and CD20, the staining appears larger than the ground truth manually stained image. In FOXP3 and PD1, some false negatives are present. In *Case 2*, positive reactions are relatively concentrated throughout the whole slides, particularly in CD20, where the staining is both concentrated and strongly reactive. Our MAS method accurately pinpoints the high-intensity areas. However, within these areas, CD3, FOXP3, and PD1 show some false positives, and CD20 reveals slight discontinuity along the patch edges. In *Case 3*, the manually stained ground truth already contains staining errors, displaying large patches of false positives. In this case, the predicted image could not avoid this false positive interference. In *Case 4*, negative reactions are predominant, with positive cells being quite sparse. Our MAS method can locate most of the positive cells among the numerous negative cells, but some errors and omissions still occur. Briefly, although there are some unideal results in special cases, our method generally identifies most positive cells, and the results remain within an acceptable range.

The input to our MAS system consists of AF and DAPI slides. The former captures a wide range of complex signal sources, including various tissue components with endogenous fluorescence such as extracellular matrix proteins, red blood cells and macrophages. However, single AF images are prone to inherent instability and noise. DAPI delivers stable and precise nuclear information, ensuring exact cell nuclei localisation and enabling refined staining impacts.[39] Therefore, we incorporated DAPI and AF as inputs to enhance the network's sensitivity, particularly for small cells. During the experiments, we compared the performance of predicted virtual mIF slides using different input modalities (Fig. 6b, Figure S1, and Table S1), including single DAPI input, single AF input, and multimodality inputs combining DAPI and AF. As illustrated in Figure S1, the results demonstrated that using AF input alone outperformed using DAPI alone due to its ability to provide a more diverse range of information from cells and tissues. However, relying solely on either DAPI or AF had limitations in accurately predicting the staining. By integrating both DAPI and AF as multimodal inputs, the performance was significantly improved, resulting in enhanced similarity between the generated predictions and the corresponding manually stained images across multiple biomarkers. These observations are further supported by the boxplots in Fig. 6b, highlighting the advantage of using multimodalities to achieve accurate and reliable staining. Looking ahead, we aim to design improved networks that can incorporate precise and stable nuclear information from DAPI through pre-training. This advancement will enable our system to generate real-time mIF images directly from AF inputs during the testing phase, without any aforehand staining.

In architecting the MAS system's model structure, we factored in the potential interdependence among different staining labels, recognising that multi-task learning can offer valuable complementary information across tasks. Generally, the shallow-level features can capture more generalised characteristics, while deep-level features are typically more specific.[40] To ensure the independence of prediction for each biomarker in the multiplex fluorescence staining task while leveraging their complementary information, we utilise a shared structure in the encoder and feature combination components, followed by parallel multi-attention decoders (Fig. 2) to generate distinct staining predictions. The comparisons in Figure S2, Figure S3, and Table 1 demonstrate that the proposed joint training strategy significantly improves PSNR, SSIM, and MI scores, highlighting its effectiveness in leveraging complementary information. Comparative experiments

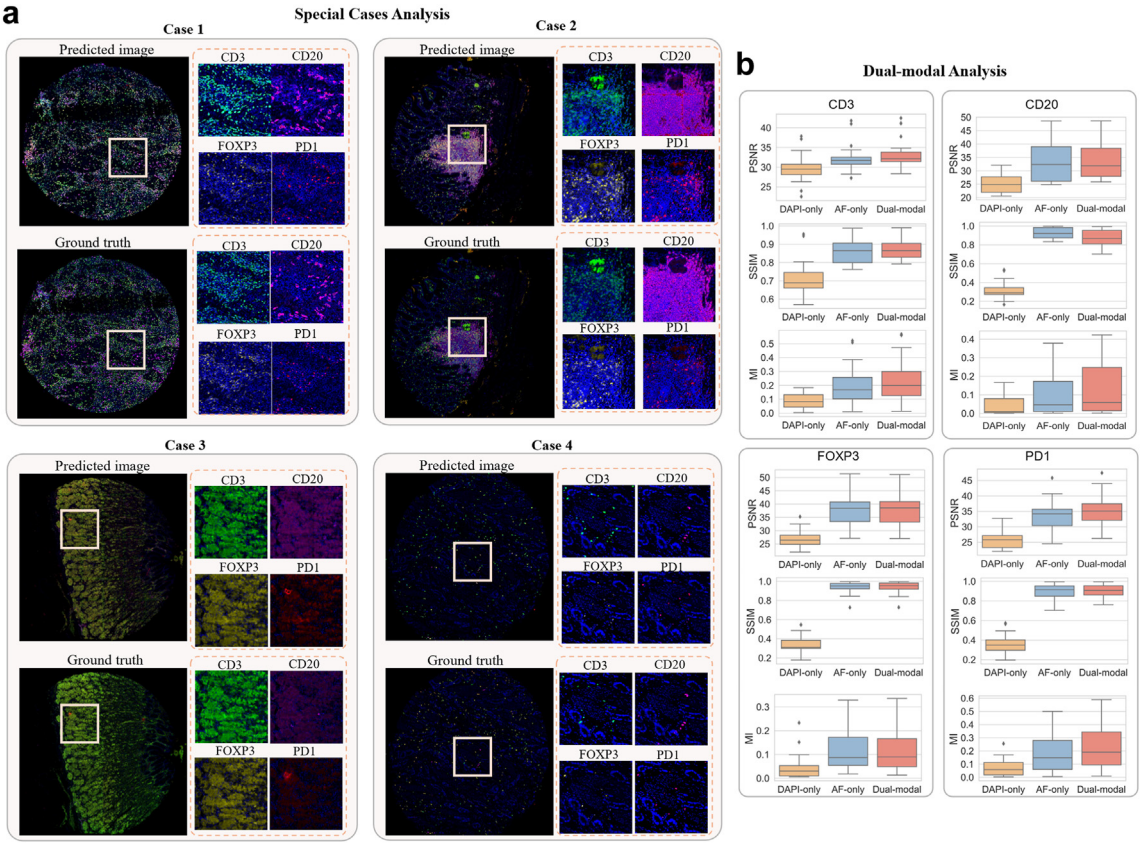


Fig. 6: **a**, Special case analysis. **b**, Boxplots of PSNR, SSIM and MI. The boxplots illustrate the PSNR, SSIM and MI values for predictions generated by dual-modal inputs and the compared single modality inputs.

against the vanilla U-Net,[41] U-Net with a residual finetune module (ReU-Net), and a modified attention U-Net[33] with a residual finetune module (Att-ReU-Net) were executed to assess each component's role. The implementation details of our method are given in Supplementary Materials. The outcomes, exemplified in Figure S4, reveal that while the virtual mIF image created by the vanilla U-Net resembles the conventional manually stained image, it exhibits a notable loss of details. Although ReU-Net and Att-ReU-Net presented improvements, they showcased false negatives. In contrast, the MAS model achieves superior staining performance with more accurate staining of true positive cells, validated by the highest PSNR and SSIM values, as depicted in Figure S5 and Table S2. In our preliminary experiments, we found that GANs were challenging to train, while CNNs achieved convergence more easily for our specific task. Figure S6 shows a sample of real-time predictions during GAN training. The GAN training process was unstable, with FOXP3 and PD1 exhibiting gradient vanishing issues. Additional comparisons between GANs and our MAS method are reported in Figure S7 and Table S3. The predictions generated by MAS surpassed those of GANs. We investigated the reasons behind the poor performance of GANs in our task. Immunofluorescence staining images often contain sparse positively stained cells, with some images appearing nearly black. This

	CD3	CD20	FOXP3	PD1	PANCK
MAS with separate training and individual biomarker prediction					
PSNR	30.86 ± 3.11	31.10 ± 6.55	36.56 ± 4.78	30.86 ± 3.11	36.83 ± 6.91
SSIM	0.56 ± 0.11	0.86 ± 0.08	0.91 ± 0.06	0.56 ± 0.11	0.85 ± 0.08
MI	0.14 ± 0.11	0.09 ± 0.09	0.08 ± 0.07	0.14 ± 0.11	0.02 ± 0.05
MAS with joint training and prediction for four biomarkers (Our approach)					
PSNR	**33.37 ± 4.03**	**33.67 ± 6.75**	**37.61 ± 5.53**	**35.09 ± 4.96**	–
SSIM	**0.87 ± 0.06**	0.87 ± 0.08	**0.94 ± 0.06**	**0.90 ± 0.06**	–
MI	**0.23 ± 0.16**	**0.14 ± 0.14**	0.12 ± 0.10	**0.24 ± 0.18**	–
MAS with joint training and prediction for five biomarkers					
PSNR	32.39 ± 4.53	33.12 ± 7.87	37.35 ± 5.17	33.38 ± 5.33	38.16 ± 8.99
SSIM	**0.87 ± 0.06**	**0.92 ± 0.06**	0.93 ± 0.05	0.89 ± 0.08	0.91 ± 0.09
MI	0.18 ± 0.14	0.10 ± 0.10	**0.13 ± 0.12**	0.16 ± 0.15	0.03 ± 0.06

The highest PSNR, SSIM, and MI results for the CD3, CD20, FOXP3, and PD1 biomarkers in the virtual mIF staining methods are highlighted in bold in Table 1. Please note that due to the suboptimal results for PANCK, no PANCK staining results are bolded.

Table 1: **Quantitative comparison of experiments for MAS in predicting different combinations of biomarkers.**

makes it difficult for the discriminator to differentiate between real and fake data, hindering effective adversarial training due to insufficient direct competition between the generator and discriminator. This observation led us to choose CNNs over GANs.

The MAS system employs a transfer learning strategy to ensure robust generalisability and swift adaptation to new tasks, particularly predicting unseen biomarkers (see Figure S8). Beyond the original four-biomarker prediction task, we also explored various biomarker combinations in transfer learning. Preliminary studies indicated that a virtual staining network for CD68 and CD66B were more challenging to achieve satisfactory performance compared to CD8, CD163, and PD-L1. The transfer learning strategy underscores the system's efficacy in realising excellent staining results for a variety of unseen biomarkers, and even adaptable for other pathological staining task (predicting DAPI, anti-MAP2, and anti-lslet1 from transmitted-light imaging,[29] see Figure S9), showcasing its adaptability through transfer learning.

Despite the promising prospects of generative AI in natural image generation tasks, its applicability in medical scenarios, where AI-generated results impact patient safety, invites scrutiny and demands rigorous usability assessments to ensure secure and effective implementation. Our proposed generative AI model underwent a thorough examination of its clinical applicability, involving a patch-based clinical observation assessment and a survival prediction analysis. The comprehensive clinical appraisal affirmed that our MAS system reliably generates mIF images with clinically significant attributes. While the prevailing standard for evaluating tumour-infiltrating immune cells hinges on intricate, labour-intensive histopathological staining methods, our model predicts survival rates and treatment responses in gastric cancer directly from autofluorescence and DAPI images, sidestepping the need for manual staining. Interestingly, in some instances, AI-predicted mIF stained slides surpassed manually stained counterparts in performance (see Figure S10), reducing extensive false-positive errors in certain cases. Attributable to the occasionally inaccurate signals in manually stained slides from varied factors like excessive antibody concentrations or suboptimal epitope retrieval methods,[42] generative AI models, bolstered by abundant training data, minimise such inaccuracies, emphasising correct protocols and representing a substantial advancement in pathological functional image generation. This underscores generative AI's potential to mitigate limitations inherent in manual staining techniques.

Our study showcases the prognostic value of predicted mIF images, offering efficiency and consistent performance while providing results comparable to standard mIF staining, and illuminates new insights into molecular-based cancer classifications through comprehensive research on tumour-infiltrating immune cells.[5,35,36,43–46] The quantification of diverse immune cell subpopulations within the tumour microenvironment reveals associations with prognosis and/or response to chemotherapy and immunotherapy, including immune checkpoint inhibitors.[5,35,36,43,45,46] Specifically, in gastric cancer, infiltration of CD3+/CD8+ T lymphocytes correlates with extended survival, whereas a high prevalence of CD163+ macrophages indicates a poorer prognosis.[35,36,44,47] These observations underscore the pivotal role of immune cell biomarkers in gastric cancer prognosis and spotlight mIF staining's potential for precise biomarker assessment. Our method generates predictive mIF images, capturing immune cell subpopulation distribution and aiding in identifying prognostic factors, thereby facilitating informed clinical decisions-making.

Previous research underscores the importance of immune cell scoring in guiding chemotherapy and immunotherapy decisions for gastric cancer, with studies substantiating the significance of this approach.[5,35,48–50] Additionally, the expression level of PD-L1 has emerged as a pivotal clinical biomarker, widely used to predict immunotherapy responses in patients with cancer. However, the clinical implementation of both immune cell and PD-L1 scoring methods is hindered by their reliance on time-consuming and costly pathological staining procedures. The proposed approach overcomes these limitations. Our generative AI model streamlines the process, enabling rapid and cost-effective generation of multiplexed staining images, including six immune cell biomarkers and PD-L1. Thus, our research introduces a viable solution that addresses the challenges associated with tumour-infiltrating immune cell and PD-L1 scoring, with significant reduced time and costs, leading to improved clinical decision-making in cancer treatment.

It is useful to acknowledge the limitations of our proposed method. In this current study, we evaluated the framework of the MAS system by predicting seven representative biomarkers. To fully assess the wide-ranging applicability of our method, it is crucial to examine its predictive efficacy across a more extensive array of biomarkers. However, due to the challenges associated with data collection, we have not yet expanded our experimental validation to include a larger set of biomarkers. Besides, future external validation is crucial to confirm the MAS system's reliability and generalisability across diverse clinical settings. By testing the model with independent datasets from various institutions, potential biases and limitations inherent to our initial study can be identified and addressed, ensuring broader applicability and enhancing clinical trust in the MAS system.

Furthermore, while the MAS system has yielded satisfactory results, we are actively working towards identifying a solution that can streamline the workflow and make our method more accessible and user-

friendly. We recognise the need to address these practical aspects to enhance the usability and adoption of the MAS system in real-world settings. Moreover, there are several valuable tasks that can be accomplished in future work. For instance, the significant time and cost associated with traditional mIF techniques present a substantial barrier, limiting their effective utilisation in investigating the diagnostic and prognostic implications of various biomarkers. In this regard, the proposed virtual mIF staining method offers a compelling solution to overcome this challenge, enabling researchers to explore the correlation between an extensive array of biomarkers and the diagnosis or prognosis of numerous diseases. Further research and development efforts are needed to fully leverage the potential of this approach and extend its applications in clinical and research settings. Moreover, different wavelengths in AF imaging can capture varied information, which might influence the efficacy of virtual staining. In future work, we will explore the correlation between different wavelength selections and staining performance for various biomarkers. This research could provide valuable insights and potentially enhance the accuracy and effectiveness of our virtual staining techniques.

In summary, deep learning-based generative AI presents a promising solution to overcome the challenges of existing manual mIF staining techniques by enabling the generation of virtual mIF staining from non-antibody-stained fluorescence imaging. Our MAS system, a dedicated framework for virtual mIF staining, achieve this by leveraging an enhanced feature extractor to derive insights from dual-modal non-antibody-stained fluorescence images. Through incorporation of self- and multi-attention mechanisms, the MAS system effectively discerns the most relevant information, facilitating rapid and reliable generation of multiplexed staining. The proposed technique brings about a significant reduction in the cost associated with mIF staining and greatly improves the clinical workflow. Furthermore, future integration of this methodology into clinical practice should empower healthcare professionals to make more informed treatment decisions, ultimately advancing precision medicine and enhancing patient care.

Contributors

ZZ, YJ and LX conceived and designed the study; YJ, ZS, TZ, WF and GL acquired the data and did the clinical assessment; ZZ, YJ, ZS, TZ did the statistical analyses; ZZ developed, trained, and applied the artificial neural network. RL and LX implemented quality control of data and the algorithms; All authors had access to the data presented in the manuscript. All authors analysed and interpreted the data; ZZ and YJ prepared the first draft of the manuscript; LX revised the manuscript. All authors read and approved the final version of the manuscript.

Data sharing statement

The datasets will be available upon reasonable request emailed to the corresponding author (lei@stanford.edu). The code for the model development is available online (https://github.com/Eva0720/MAS).

Declaration of interests

All authors declare no competing interests.

Acknowledgements

This work is supported by the Stanford 2022 HAI Seed Grant and National Institutes of Health 1R01CA256890.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2024.105287.

References

1 Andreou C, Weissleder R, Kircher MF. Multiplexed imaging in oncology. *Nat Biomed Eng*. 2022;6:527–540.
2 Schürch CM, Bhate SS, Barlow GL, et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell*. 2020;182:1341–1359.e19.
3 Snyder MP, Lin S, Posgai A, et al. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature*. 2019;574:187–192.
4 Bayer-Garner IB, Korourian S. Plasma cells in chronic endometritis are easily identified when stained with syndecan-1. *Mod Pathol*. 2001;14:877–879.
5 Foersch S, Glasner C, Woerl A-C, et al. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. *Nat Med*. 2023;29:430–439.
6 Rizzardi AE, Johnson AT, Vogel RI, et al. Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagn Pathol*. 2012;7:42.
7 Angelo M, Bendall SC, Finck R, et al. Multiplexed ion beam imaging of human breast tumors. *Nat Med*. 2014;20:436–442.
8 Black S, Phillips D, Hickey JW, et al. CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat Protoc*. 2021;16:3802–3835.
9 Duraiyan J, Govindarajan R, Kaliyappan K, Palanisamy M. Applications of immunohistochemistry. *J Pharm Bioallied Sci*. 2012;4:S307–S309.
10 Im K, Mareninov S, Diaz MFP, Yong WH. An introduction to performing immunofluorescence staining. *Methods Mol Biol*. 2019;1897:299–311.
11 Armanious K, Jiang C, Fischer M, et al. MedGAN: medical image translation using GANs. *Comput Med Imaging Graph*. 2020;79:101684.
12 Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-Image translation with conditional adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017:1125–1134.
13 Xing L, Giger ML, Min JK. *Artificial intelligence in medicine: technical basis and clinical applications*. Academic Press; 2020.
14 Zhou Z, Wang Y, Yu J, Guo Y, Guo W, Qi Y. High spatial–temporal resolution reconstruction of plane-wave ultrasound images with a multichannel multiscale convolutional neural network. *IEEE Trans Ultrason Ferroelectr Freq Control*. 2018;65:1983–1996.
15 Zhou Z, Guo Y, Wang Y. Handheld ultrasound video high-quality reconstruction using a low-rank representation multipathway generative adversarial network. *IEEE Trans Neural Netw Learn Syst*. 2021;32:575–588.
16 Bai B, Yang X, Li Y, Zhang Y, Pillar N, Ozcan A. Deep learning-enabled virtual histological staining of biological samples. *Light Sci Appl*. 2023;12:57.
17 Cao R, Nelson SD, Davis S, et al. Label-free intraoperative histology of bone tissue via deep-learning-assisted ultraviolet photoacoustic microscopy. *Nat Biomed Eng*. 2023;7:124–134.
18 Rivenson Y, Wang H, Wei Z, et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng*. 2019;3:466–477.
19 Ghahremani P, Li Y, Kaufman A, et al. Deep learning-inferred multiplex immunofluorescence for immunohistochemical image quantification. *Nat Mach Intell*. 2022;4:401–412.
20 Wang H, Rivenson Y, Jin Y, et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat Methods*. 2019;16:103–110.
21 Wu Y, Luo Y, Chaudhari G, et al. Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram. *Light Sci Appl*. 2019;8:25.

22 Li D, Hui H, Zhang Y, et al. Deep learning for virtual histological staining of bright-field microscopic images of unlabeled carotid artery tissue. *Mol Imaging Biol*. 2020;22:1301–1309.
23 Zhang Y, de Haan K, Rivenson Y, Li J, Delis A, Ozcan A. Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. *Light Sci Appl*. 2020;9:78.
24 Jo Y, Cho H, Park WS, et al. Label-free multiplexed microtomography of endogenous subcellular dynamics using generalizable deep learning. *Nat Cell Biol*. 2021;23:1329–1337.
25 Sun K, Chen Z, Wang G, Liu J, Ye X, Jiang Y-G. Bi-Directional feature fusion generative adversarial network for ultra-high resolution pathological image virtual Re-staining. In: *2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. 2023:3904–3913.
26 He B, Bukhari S, Fox E, et al. AI-enabled in silico immunohistochemical characterization for Alzheimer's disease. *Cell Reports Methods*. 2022;2:100191.
27 Zhang R, Cao Y, Li Y, et al. MVFStain: multiple virtual functional stain histopathology images generation based on specific domain mapping. *Med Image Anal*. 2022;80:102520.
28 Bai B, Wang H, Li Y, et al. Label-free virtual HER2 immunohistochemical staining of breast tissue using deep learning. *BME Front*. 2022;2022:9786242.
29 Christiansen EM, Yang SJ, Ando DM, et al. In silico labeling: predicting fluorescent labels in unlabeled images. *Cell*. 2018;173:792–803.e19.
30 Wu E, Trevino AE, Wu Z, et al. 7-UP: generating in silico CODEX from a small set of immunofluorescence markers. *PNAS Nexus*. 2023;2:pgad171.
31 Taylor L, Nitschke G. Improving deep learning with generic data augmentation. In: *2018 IEEE symposium series on computational intelligence (SSCI)*. 2018:1542–1547.
32 Zhang H, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. In: *Proceedings of the 36th international conference on machine learning*. PMLR; 2019:7354–7363.
33 Oktay O, Schlemper J, Folgoc LL, et al. Attention U-net: learning where to look for the pancreas. *arXiv*; 2018. https://arxiv.org/abs/1804.03999v3. Accessed May 25, 2023.
34 Hendry S, Salgado R, Gevaert T, et al. Assessing tumor infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the international immuno-oncology biomarkers working group. *Adv Anat Pathol*. 2017;24:311–335.
35 Jiang Y, Zhang Q, Hu Y, et al. ImmunoScore signature: a prognostic and predictive tool in gastric cancer. *Ann Surg*. 2018;267:504–513.
36 Jiang Y, Xie J, Han Z, et al. Immunomarker support vector machine classifier for prediction of gastric cancer survival and adjuvant chemotherapeutic benefit. *Clin Cancer Res*. 2018;24:5574–5584.
37 Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res*. 2010;1:274–278.
38 Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
39 Otto F. Chapter 11 DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. In: Darzynkiewicz Z, Crissman HA, eds. *Methods in cell biology*. Academic Press; 1990:105–110.
40 Long M, Cao Y, Wang J, Jordan M. Learning transferable features with deep adaptation networks. In: *Proceedings of the 32nd international conference on machine learning*. PMLR; 2015:97–105.
41 Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical image computing and computer-assisted intervention – MICCAI 2015*. Cham: Springer International Publishing; 2015:234–241.
42 Zaqout S, Becker L-L, Kaindl AM. Immunofluorescence staining of paraffin sections step by step. *Front Neuroanat*. 2020;14. https://www.frontiersin.org/articles/10.3389/fnana.2020.582218. Accessed May 25, 2023.
43 Fridman WH, Zitvogel L, Sautès–Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol*. 2017;14:717–734.
44 Li C, Xu X, Wei S, Jiang P, Xue L, Wang J. Tumor-associated macrophages: potential therapeutic strategies and future prospects in cancer. *J Immunother Cancer*. 2021;9:e001341.
45 Mahajan UM, Langhoff E, Goni E, et al. Immune cell and stromal signature associated with progression-free survival of patients with resected pancreatic ductal adenocarcinoma. *Gastroenterology*. 2018;155:1625–1639.e2.
46 Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet*. 2018;391:2128–2139.
47 Rihawi K, Ricci AD, Rizzo A, et al. Tumor-associated macrophages and inflammatory microenvironment in gastric cancer: novel translational implications. *Int J Mol Sci*. 2021;22:3805.
48 Keren L, Bosse M, Marquez D, et al. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell*. 2018;174:1373–1387.e19.
49 Huang W, Jiang Y, Xiong W, et al. Noninvasive imaging of the tumor immune microenvironment correlates with response to immunotherapy in gastric cancer. *Nat Commun*. 2022;13:5095.
50 Jiang Y, Xie J, Huang W, et al. Tumor immune microenvironment and chemosensitivity signature for predicting response to chemotherapy in gastric cancer. *Cancer Immunol Res*. 2019;7:2065–2073.