

Received 28 April 2022; revised 12 June 2022; accepted 22 June 2022. Date of publication 19 July 2022; date of current version 5 August 2022. The review of this paper was arranged by Editor P. Bonato.

Digital Object Identifier 10.1109/OJEMB.2022.3192103

Identification of Tissue Types and Gene Mutations From Histopathology Images for Advancing Colorectal Cancer Biology

YUQI JIANG¹, CECILIA K. W. CHAN², RONALD C. K. CHAN³, XIN WANG¹, NATHALIE WONG¹, KA FAI TO³, SIMON S. M. NG⁴, JAMES Y. W. LAU², AND CARMEN C. Y. POON⁵ (Senior Member, IEEE)

¹Department of Surgery, The Chinese University of Hong Kong, Hong Kong SAR

²Division of Vascular and General Surgery, Department of Surgery, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong SAR

³Department of Anatomical and Cellular Pathology, The Chinese University of Hong Kong, Hong Kong SAR

⁴Division of Colorectal Surgery, Department of Surgery, The Chinese University of Hong Kong, Hong Kong SAR

⁵GMed IT Ltd., Hong Kong SAR

CORRESPONDING AUTHORS: JAMES Y. W. LAU; CARMEN C. Y. POON (e-mail: laujyw@surgery.cuhk.edu.hk; cpoon@link.cuhk.edu.hk)

The work was supported by the General Research Fund and Innovation and Technology Fund, Hong Kong SAR.

This article has supplementary downloadable material available at <https://doi.org/10.1109/OJEMB.2022.3192103>, provided by the authors.

ABSTRACT *Objective:* Colorectal cancer (CRC) patients respond differently to treatments and are subclassified by different approaches. We evaluated a deep learning model, which adopted endoscopic knowledge learnt from AI-doscopist, to characterise CRC patients by histopathological features. *Results:* Data of 461 patients were collected from TCGA-COAD database. The proposed framework was able to 1) differentiate tumour from normal tissues with an Area Under Receiver Operating Characteristic curve (AUROC) of 0.97; 2) identify certain gene mutations (MYH9, TP53) with an AUROC > 0.75; 3) classify CMS2 and CMS4 better than the other subtypes; and 4) demonstrate the generalizability of predicting KRAS mutants in an external cohort. *Conclusions:* Artificial intelligent can be used for on-site patient classification. Although KRAS mutants were commonly associated with therapeutic resistance and poor prognosis, subjects with predicted KRAS mutants in this study have a higher survival rate in 30 months after diagnoses.

INDEX TERMS AI-doscopist, medical device, deep learning, tumour heterogeneity, precision medicine..

IMPACT STATEMENT In external validation, subjects identified with KRAS mutants by an artificial intelligent system trained by both endoscopy and histopathology features have a higher survival rate in 30 months after diagnoses.

I. INTRODUCTION

Colorectal cancer (CRC) remains to be one of the commonest cancers worldwide [1]. An AI-doscopist (a.k.a. Artificial Intelligent Endoscopist) can facilitate endoscopists to improve polyp detection rate during colonoscopy (please refer to Supplementary Materials, Fig. S1) [2]. Nevertheless, this artificial intelligent (AI) system did not perform well in characterizing the polyp subtypes at present. As shown in Fig. 1, suspicious tissues have to be resect for subtype classification in order to determine an appropriate therapy and follow up plan [3], [4]. On the other hand, histopathology features have been used for tumour stratification. Deep

learning has also been used to classify histopathology images of breast cancer [5], lung cancer [6], and pan-cancer [7]–[9], as well as to classify Consensus Molecular Subtypes (CMS) and predict prognosis of CRC patients [10], [11]. Nevertheless, to our best knowledge, no study has investigated a deep learning model that utilizes information from both endoscopy and histopathology, which is the current clinical practice.

In this study, we adopted the endoscopic knowledge of AI-doscopist to further build a deep learning framework that consists of several models with the same backbone network. The new framework was designed for

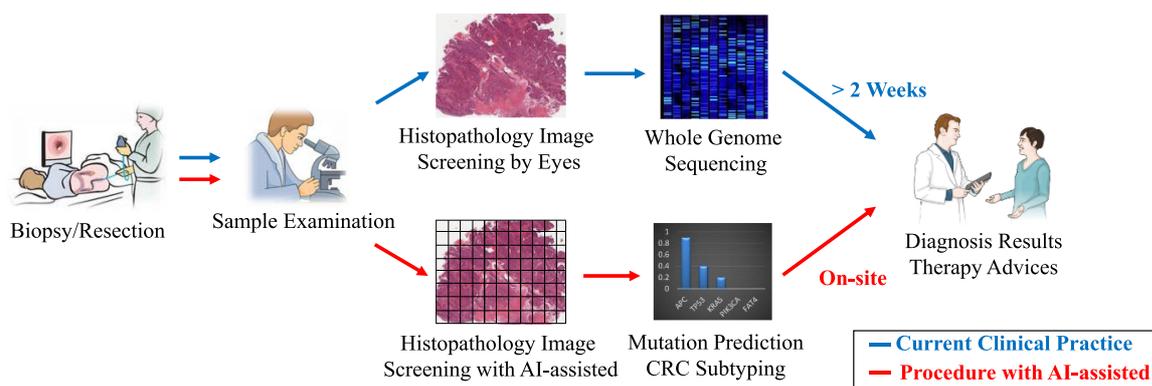


FIGURE 1. An illustration of the usage of the intra-operative system in future. Biopsy is currently needed for diagnosing colorectal cancer. Nevertheless, the accuracy of the diagnosis depends on the skills and experience of the clinicians in selecting the site of biopsy, as well as the experience and knowledge of the pathologists in preparing the tissue and determining the cancer subtype. Although next generation sequencing is a useful tool for obtaining the gene mutation spectrum, it is relatively expensive and not always available, especially in low-resource clinical settings. To tackle the above issues, a computer-aided system is designed to assist clinicians to screen the histopathology images, predict gene mutations of CRC and classify the subtypes of CRC.

1) differentiating CRC tissue from normal tissue; 2) identifying suspicious regions with commonly known mutated genes of CRC; and 3) classifying the subtypes of CRC. We trained these models to learn histopathologic features from The Cancer Genome Atlas-Colon Adenocarcinoma (TCGA-COAD) public database and verified their performance on both public and private databases. Regarding the last sub-task, CMS, which is a robust CRC classification standard derived from gene expression data [12], was used as the reference in this study. In brief, the four CMS and their corresponding distinct features are as follows: CMS1) Microsatellite Instability Immune (14%) - hypermutated, microsatellite unstable and strong immune activation; CMS2) Canonical (37%) - epithelial, marked WNT and MYC signalling activation; CMS3) Metabolic (13%) - epithelial and evident metabolic dysregulation; and CMS4) Mesenchymal (23%) - prominent transforming growth factor- β activation, stromal invasion and angiogenesis. Samples with mixed features (13%) possibly represent a transition phenotype or intratumoural heterogeneity.

II. RESULTS

We evaluated the deep learning models on a Nvidia GTX 1080Ti device. The resolution of the input image tiles was fixed to 512×512 . The image tiles were processed at around 26 to 30 frames per second (fps). Considering that at 5.0 magnification ratio, each slide has 94 tiles on average, the proposed models were capable of classifying tissue types for one slide within one minute, which is adequate for screening flash-frozen slide resected and prepared during surgical operation. Fig. 2 illustrates the typical outputs of the three sub-tasks from our model.

A. BINARY CLASSIFICATION OF PRIMARY TUMOUR FROM SOLID TISSUE NORMAL IN EACH TILE AND SLIDE

Fig. 3 presents the Area Under Receiver Operating Characteristic curves (AUROC) of the model for classifying primary

tumour (PT) from solid tissue normal (STN) in the validation and testing datasets. When evaluated on the validation dataset, the performance of each sub-classifier varied from 0.88 to 0.97. When evaluated tile-by-tile on the testing dataset, the ensemble model yields an AUROC of 0.97 (95% confidence interval [CI_{95%}], 0.93–0.99). Furthermore, the ensemble model can correctly identify 127 out of 131 PT slides (sensitivity = 0.969) and 19 out of 20 STN slides (specificity = 0.950).

B. IDENTIFICATION OF SELECTED GENE MUTATIONS IN EACH TILE

Fig. 4 shows the AUROC of the identification of gene mutations. The AUROCs of all 13 selected genes were above 0.64. At 5.0 magnification ratio (MR), the top 3 mutated genes those can be most accurately identified were MYH9 with AUROC of 0.785 (CI_{95%}, 0.776–0.800), NIN with AUROC of 0.739 (CI_{95%}, 0.727–0.745) and TP53 with AUROC of 0.735 (CI_{95%}, 0.726–0.746). On the other hand, when the MR were increased to 20.0, MTOR with AUROC of 0.791 (CI_{95%}, 0.785–0.811), MYH9 with AUROC of 0.786 (CI_{95%}, 0.778–0.798) and TP53 with AUROC of 0.741 (CI_{95%}, 0.735–0.748) were the top 3 genes whose mutation can be most accurately identified. Furthermore, as shown in Fig. 4(b), mutation in the APC and TP53 genes can be identified with a sensitivity of over 90% when operated at a specificity of 30%. The findings were further confirmed by the results trained and tested on images with a 20.0 MR.

C. CLASSIFICATION OF THE CONSENSUS MOLECULAR SUBTYPES IN EACH TILE AND SLIDE.

The results of the tile-based and slide-based classification of the CMS at 5.0 and 20.0 MR was illustrated as the confusion matrices in Fig. S2(a)–(d). CMS2 were classified with a sensitivity and specificity of (0.591, 0.730) and (0.727, 0.646) for slides at 5.0 and 20.0 MR, respectively. The sensitivity of

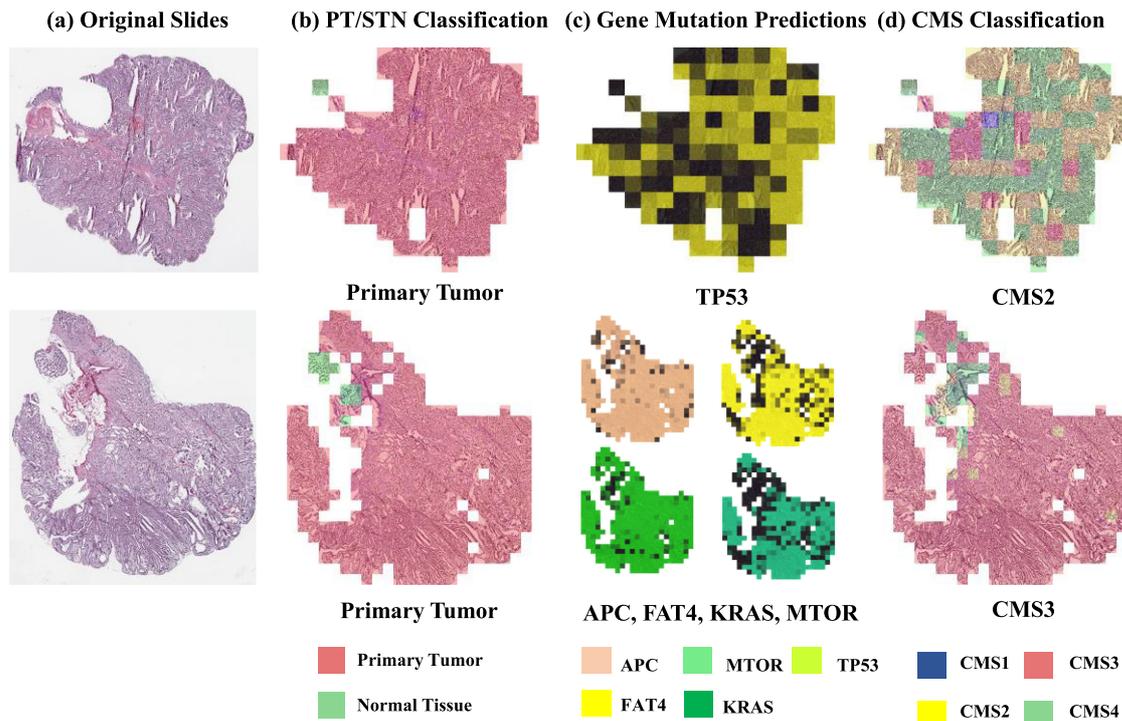


FIGURE 2. Two typical examples of the proposed model in performing the three sub-tasks in this study. Histopathology tiles at 5.0 magnification ratio were used as inputs to the model and the output from each sub-task were overlaid on the (a) Original slide. (b) Each tile was independently marked as either primary tumor or normal tissue. A majority voting of all tiles was used to determine the decision for each slide. (c) The mutation of each selected gene was determined for each tile. Respect to the groundtruth, only specific mutated genes' heatmaps were displayed for each slide. (d) A CMS was determined for each tile and the majority voting of all tiles was used to determine the decision for each slide. In some slides, the intratumoural heterogeneity can be clearly observed.

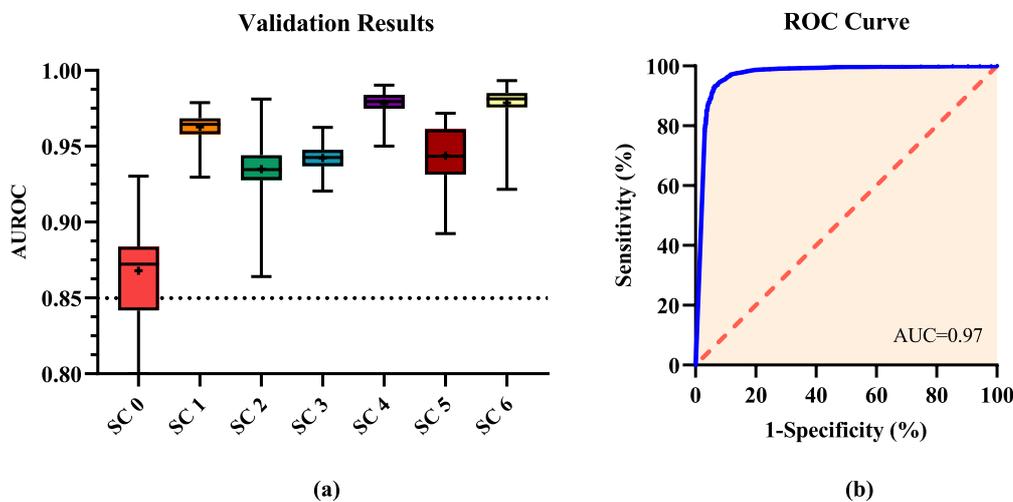


FIGURE 3. The tile-based performance of the binary classification of primary tumor from solid tissue normal by ensemble learning. (a) The mean and standard deviation of the area under receiver operating characteristic (AUROC) curves of each sub-classifier evaluated on the validation dataset for different epochs. (b) The ROC curve of the ensemble model evaluated on the testing dataset.

classifying CMS1 and CMS4 were less than 0.5 using slides at 5.0 and 20.0 MR. CMS3 was the most difficult class to identify. Fig. S2(e) shows the Sankey diagram, which indicates the portion of slides that the AI-classifier mapped to the RF-classifier on 5.0 magnified slides.

D. INDEPENDENT VALIDATION ON KRAS MUTANTS FOR AN EXTERNAL COHORT IN EACH TILE AND SLIDE

The models were further evaluated on independent CRC cohorts comprising formalin-fixed paraffin embedding (FFPE) slides ($n = 40$) from Prince of Wales Hospital. Those slides

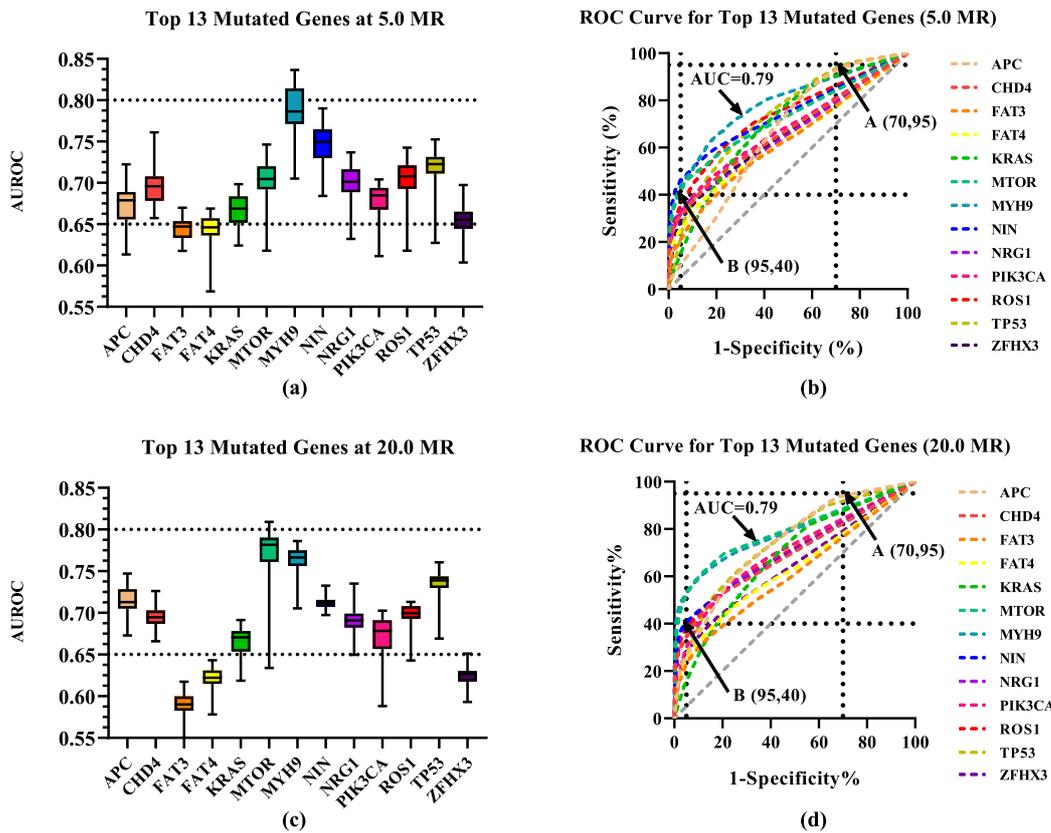


FIGURE 4. The AUROC curves for the identification of the mutations in 13 selected genes from the tiles of the histopathology images with (a)–(b) 5.0 and (c)–(d) 20.0 magnification ratios. MYH9 and TP53 were the two genes that can be consistently identified in images of both 5.0 and 20.0 MR. The ROC curves further demonstrated that the identification can be operated at a relatively high sensitivity (>90%) for a specificity of 30%.

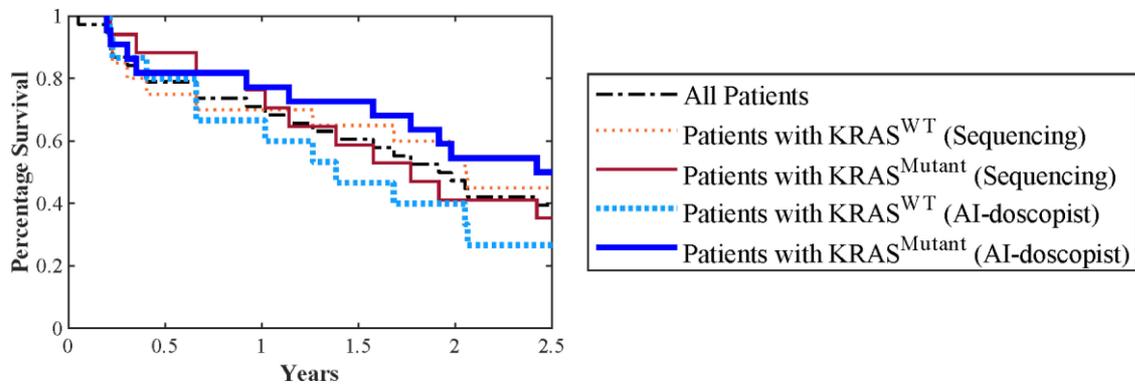


FIGURE 5. Survival curves for KRAS-wild type and KRAS-mutant patients divided into two groups using (a) sequencing data as reference, and (b) the prediction of KRAS gene mutations of the proposed model.

were confirmed as tumour samples and sequenced for KRAS mutations. One slide was excluded for KRAS mutation identification because of insufficient tissue for obtaining the ground truth label for sequencing. Predictions were made independently on a total of 1,493 tiles at the 5.0 MR. A final decision was made for each slide and compared with the result obtained from the sequencing method. The ensemble model classified 39 out of 40 tumour slides (sensitivity = 0.975) at 5.0 MR. KRAS mutations were identified with AUROC of

0.594 (CI_{95%}, 0.568-0.619) and 0.606 (CI_{95%}, 0.582-0.631) at 5.0 and 20.0 MR, respectively. On slide level, 13 out of 17 slides (sensitivity = 0.765) with KRAS mutants were identified and 13 out of 22 slides (specificity = 0.591) with KRAS wild types were detected. Representative tiles were selected as shown in Supplementary Materials, Fig. S3 to indicate the difference between KRAS mutants and wild types. Furthermore, Since the subjects were approximately diagnosed on the same date, the Kaplan-Meier curves shown in Fig. 5 were plotted

to indicate the fraction of patients who survive after 1st Jan. 2019 in each subgroup of subjects.

III. DISCUSSION

AI has gained substantial interests in the healthcare domain, including using deep learning to associate the histopathology features with genomic, transcriptomic and biomarkers for multiple cancer types. Our study contributed uniquely in the following aspects: 1) we proposed a unique, unified framework to perform three sub-tasks related to gene mutation identification and subtyping of CRC, using pre-trained weights from AI-doscopist, i.e., deep features that learnt to localize colorectal polyps in endoscopy videos [2]. 2) The proposed deep learning model was able to differentiate tumour versus normal tissue with high accuracy (0.97 AUC) in real-time, which is potentially beneficial to screening histopathology slides during surgical operation. 3) The correlation between phenotype and genotype of CRC were confirmed and spatially visualized by activation map generated by our model, and histopathology features are associated with certain gene alterations, i.e., APC, MTOR, MYH9, NIN, and TP53 (please refer to examples in the Supplementary Table S-I). 4) The external validation of our model on an independent cohort demonstrated the feasibility and generalizability of our model applied in real clinical practice. Ultimately, the model can be integrated with a microscopic system for cancer subtyping and identification of gene mutations in real-time. For example, by providing instance therapeutic advice to clinicians regarding the site of biopsy, whether a resection margin is clear, and/or whether certain type of targeted therapy should be avoided.

Our results indicate that using the proposed framework, PT can be differentiated from STN with both the sensitivity and specificity of >95%, either at the tile or slide levels on TCGA-COAD database. Similarly, our model achieved a sensitivity of 0.974 in identifying tumour slides on external cohorts in the same settings as those on public database. Certain gene mutations (MYH9 and TP53) can be identified consistently in the tiles with AUROC close to 0.8, either on tiles with 5.0 or 20.0 MR. KRAS mutations can be identified with AUROC around 0.66 on public database and around 0.6 on collected database. Nevertheless, around 50% of the CMS outputted by the AI-classifier were different from that of the RF-classifier. CMS2 was the subtype that agreed most between the AI-classifier and the RF-classifier, with a sensitivity and specificity of 0.727 and 0.646, respectively.

Our study showed that even trained with a weak label, i.e., all tiles in a slide were labelled with the same set of gene mutations, mutations for a selected panel of genes can still be identified with promising accuracy. Seven genes (CHD4, MTOR, MYH9, NIN, PIK3CA, TP53, ROS1) can be identified from histopathology images at both 5.0 and 20.0 MR for an AUROC of over 0.7. Our results indicated that increasing the inspection of histopathology slides from 5.0 to 20.0 MR did not significantly improve the identification accuracy of gene mutations on average. Rather, there were only specific genes

(APC, MTOR) benefited from circumstantial histopathology features. In the independent cohort, our model demonstrated good performance on tissue differentiation and a moderate drop (0.06) on KRAS mutations identification, which possibly owed to several factors: 1) slides are acquired in a heterogeneous way, including scanner modalities, slide preparation processes etc; 2) we trained our model on flash-frozen slides yet validated on FFPE slides; 3) the representation complexity of phenotype among different cohorts may varies extensively. Overall, the study showed promising results to be adopted in real clinical settings. Amongst the different gene mutations being studied, patients with KRAS mutants are resistant to anti-EGFR therapy. TP53 is a gene that is related to the activation of cellular stress and exerts multiple, anti-proliferative functions. TP53 controls genes that involve in cell-cycle inhibition, apoptosis, maintaining genetic stability, and inhibit blood-vessel formation [13]. Mutation in TP53 is also found in other types of cancer, such as breast cancer, bone and soft tissue sarcomas, brain tumours adrenocortical carcinomas (ADC), pancreatic adenocarcinoma, and prostate cancer [14]–[17]. Patients with a TP53-mutated breast cancer were found to have a poor prognosis in certain type of treatment [18]. On the other hand, MYH9 encodes the 224-kD nonmuscle myosin heavy chain IIA (MYHIIA) polypeptide, which is present in platelets, monocytes and granulocytes for functions such as cytokinesis, cell motility and maintenance of cell shape [19]. MYH9 is more commonly studied in nondiabetic end-stage renal disease [20] and also found in some types of breast cancer [21]. The Supplementary Table S-II summarizes the major current understanding of the 13 selected genes in other types of cancers and diseases. The identification accuracy of gene mutations from histopathology images were comparable to those for non-small cell lung cancer [6] and better than other study for CRC [9], which shared some common gene mutations such as KRAS and TP53. The proposed deep learning model can generate the activation map (also known as heatmap) for the visualization of certain gene alterations, which links the genomic and molecular traits to specific histopathological features. On the other hand, as shown in Fig. 4(b) and (d), Point A illustrates an operating point when the model can be used to accurately identify regions with gene mutations of MYH9 and TP53, while Point B is an operating point that can confirm regions without mutation of genes such as MTOR, MYH9, and NIN. This can be beneficial to tailored treatment plans for the patients instead of performing molecular profiling by transcriptomic sequencing.

Molecular stratification of cancer based on histopathological features has not been fully explored. A recent study [10] attempted to classify CMS of CRC from hematoxylin and eosin-stained slides by deep learning. By training on the FOCUS dataset, a system named imCMS was built to classify CMS with an average AUROC of 0.84 on the TCGA database [10]. The CMS labels were derived from a random forest (RF)-classifier that learnt from a genetic spectrum of 20 thousand mutated genes [22]. Our AI-classifier was designed to learn 13 selected gene mutations and classify the CMS based on the

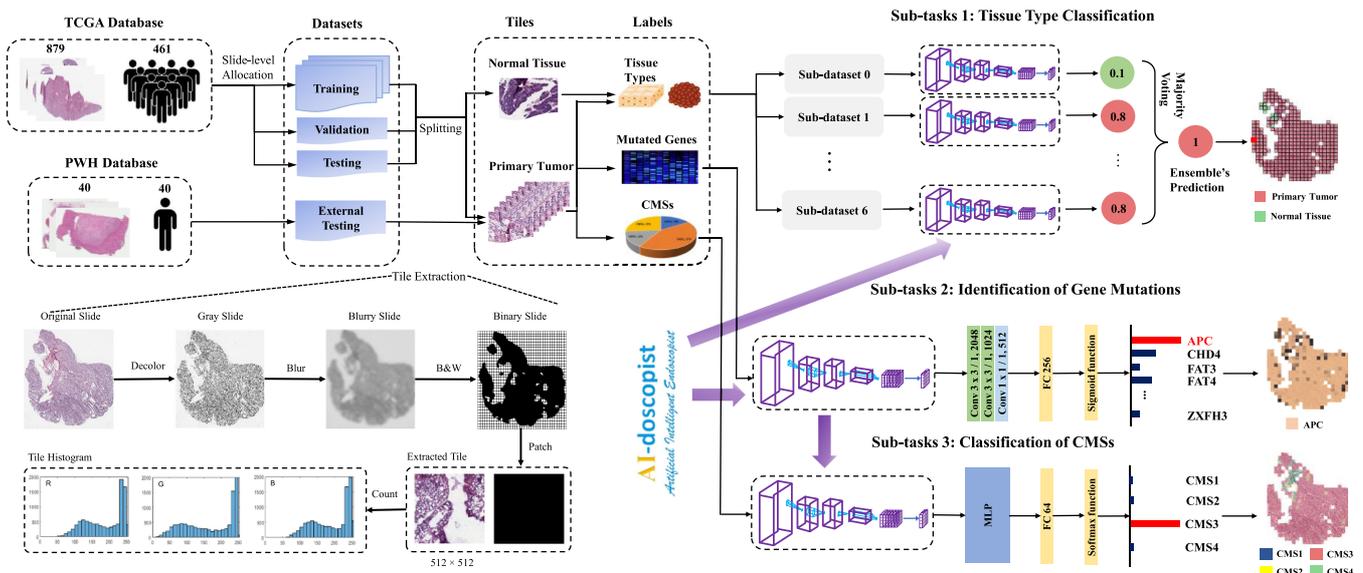


FIGURE 6. An overview of the study design, where three sub-tasks were conducted by a deep learning model that accepts histopathology imaging tiles as input. (a) A set of whole slide images for colorectal cancer were downloaded from the TCGA-COAD database and collected from PWH database. The slides were split into the training, validation and testing datasets. Each slide was split into tiles of two magnification ratios, i.e., 5.0 and 20.0. (b) The first sub-task aims to differentiate primary tumour from solid tissue normal of each tile by ensemble learning. The decision on a slide was made by the majority voting of all tiles from that slide. (c) The second sub-task aims to detect whether a subset of genes was mutated in each tile of a slide. (d) The third sub-task aims to classify a slide into one of the four CMSs. (e) The tiles are extracted by using a series of image processing algorithms to keep the informative ones.

selected gene mutations. Although analysis across the complete transcriptome and the functional spectrum of the genes can reduce the disagreement between different classifiers [22], transcriptomic analysis is considerably expensive and may not be affordable for most patients. CMS1 was mainly driven by mutations in genes which were not in our selected gene list [3]. CMS2, CMS3 and CMS4 were mainly driven by mutations in APC, KRAS, TP53, SMAD4, and PIK3CA, but CMS3 and CMS4 involved also other genomic alterations and cellular process dysregulation. This partly explains why CMS2 and CMS4 can be identified more readily than the other two subtypes from the histopathologic features of the TCGA database.

IV. CONCLUSION

The findings of this study support the application of an artificial intelligent system for real-time classification of tissue types, as well as the identification of certain gene alterations and CMS from histopathological features. Moreover, the study elucidates the association between genotype and phenotype and explores intra-tumour heterogeneity by a spatial visible approach. Although KRAS mutants is commonly associated with therapeutic resistance and poor prognosis, this study showed that subjects with predicted KRAS mutants have a higher survival rate in two and a half years after diagnosis.

V. MATERIALS AND METHODS

A. STUDY PROTOCOL

The complete workflow of this study is summarized in Fig. 6. Whole slide images were downloaded and randomly allocated

into three datasets for training, validating, and testing the deep learning model, respectively. Each slide was further split into tiles at two MRs, i.e., 5.0 and 20.0. Deep learning models were designed to complete three different tasks: 1) to differentiate PT from STN of CRC; 2) to predict the mutations of a selected subset of genes of CRC in whole-slide imaging tiles; and 3) to perform molecular stratification of CRC, i.e., CMS subtyping.

Based on the current clinical practice in CRC diagnosis, the proposed framework was designed by transferring the endoscopic knowledge learnt previously by AI-doscopist to the histopathological domain for further modelling. In a pilot study, we implemented several training strategies to verify the benefits of using endoscopy knowledge as the pre-trained weights. Compared to using the pre-trained weights from the ImageNet Challenge, the use of endoscopy knowledge in the initialisation process can bring about 5% improvement in both sensitivity and specificity. When combined with other training strategies, the increment can be further raised to 10%. Moreover, the training process of the redesigned framework was accelerated by around 30%.

Based on this redesigned framework, in the first sub-task, ensemble learning was used to tackle the extremely imbalance number of tumour and normal samples. Seven sub-classifiers, each having the same architecture adopted from AI-doscopist [2], were used to classify the tiles on its independent sub-dataset. The outputs of all sub-classifiers were aggregated, and a final prediction was major-voted for each tile. The decision to a slide was made by the majority voting of all informative tiles within that slide.

Sub-task 2 was designed to predict the gene mutations from the whole-slide images by multi-instance learning (MTL).

We excluded those which belong to most frequently mutated genes in all cancers, such like TTN, MUC16, and SYNE1 [23], and selected 13 mutated genes as target for identification. The proposed model was trained and validated on 515 slides and tested on 171 slides.

Sub-task 3 was designed to perform molecular stratification of CRC (i.e., CMS subtyping) from histopathological slides. Since gene mutations were believed to be associated with CMS subtyping, the weights learnt from sub-task 2 was transferred as pre-trained weights, i.e., to link up the molecular traits with the histopathology patterns.

B. ALGORITHM DESCRIPTION

The backbone network and the pre-trained weights used in this study were adopted from AI-doscopist [2]. In brief, the feature extractor was modified from ResNet50 [24], which consisted of 16 residual blocks that were built from three convolutional layers with increasing channel widths. The network was proven to have good trade-off between speed and accuracy in colorectal polyp localization and classification tasks [25]. The deep features of AI-doscopist was directly transferred to train and test the model on histopathology slides. Different output layers were modified to match the respective desired outputs of each sub-tasks. For sub-task 1, three convolution layers and one fully connected layer were added as the head of model for binary classification. For sub-task 2, the final output layer was extended to 13 nodes, corresponding to each selected gene. For sub-task 3, a multi-layer perceptron with seven fully connected layers and four dropout layers was used as the classification network.

The model was initialized with the pre-trained weights from AI-doscopist and fine-tuned by histopathology image tiles for 100 epochs, using the Adam optimization function with a learning rate of 0.001 and weight decay of 0.0006. Seven sub-classifiers, each with the same CNN architecture, was trained on different sub-datasets and their outputs were aggregated. For sub-task 3, the pre-trained weights from sub-task 2 was adopted. Histopathology images were only used to train the final MLP for 100 epochs, using the same conditions aforementioned. The learning process was guided by the validation process to avoid overfitting.

C. DATASET PREPARATION

Histopathology images and the corresponding clinical information and gene mutations were downloaded from TCGA-COAD online database. Since the system is designed to be ultimately used by surgeons to characterize specimens during operations, only the 879 flash-frozen slides from 461 patients (774 PT and 105 STN slides) were downloaded and used in the analysis. The slides were split into training, validation, and testing datasets in ratio of 70:15:15.

Each whole-slide image was split into non-overlapping tiles of 512×512 pixels. To study the effects of the field of view, two magnification ratios, i.e., 5.0 and 20.0, were selected for tile extraction. An adaptive thresholding algorithm was used to binarise the tiles. Tiles with white background area

TABLE 1. The Number of Slides and Their Corresponding Class Labels and Subtypes Collected From TCGA-COAD Database

Class ^a	Subtype	Training	Validation	Testing	Total
PT	CMS1	88	23	24	135
	CMS2	177	39	44	260
	CMS3	57	16	16	89
	CMS4	120	30	25	175
	NOLBL	73	20	22	115
STN		73	12	20	105
Total		588	140	151	879

^aPT = primary tumour; STN = solid tissue normal; CMS = consensus molecular subtypes; NOLBL = no labels.

over 70% were excluded. Then, the variance of the tiles was estimated to exclude the outliers. The tiles were used to train each sub-classifier for sub-task 1. PT tiles were used to train the model in sub-task 2 and the CMS labels were used to train the model in sub-task 3.

For the first sub-task, each slide was labelled as one single class, i.e., PT or STN. Each tile was assigned the same label as the slide, i.e., a weakly supervised learning approach was adopted. Similarly, for the second sub-task, all PT slides of the same subject were labelled with the same set of gene mutations. Different from the sequencing method, where usually only a part of the tissue was being sampled, the proposed model made predictions on gene mutations independently for each tile of each slide. Therefore, the heatmaps generated by the model did not only take into account possible gene mutations in the cancerous tissues, but could also show the conditions for the tissues around the tumour cells. Although there were over 10,000 gene mutation information for each subject in the TCGA-COAD database, only 53 gene mutations that were found in more than 10% of all TCGA-COAD cases and have the representative association between phenotype and genotype were selected for preliminary evaluation. Amongst the 53 mutated genes, 13 genes which can be predicted with higher accuracy (i.e., AUROC > 0.6) using a preliminary model were further chosen for learning and testing. For the third sub-task, CMS labels which were obtained by a random-forest (RF) machine learning algorithm based on genomic spectrum were used [12], [22]. Tables 1 and 2 provide the details of the slides and tiles used in this study.

To assess the generalizability of proposed models, 40 FFPE tumour slides from an independent CRC patient cohort were collected from our Hospital, Prince of Wales Hospital (PWH), Hong Kong SAR. Consents were waived for deceased patients and the study protocol has been approved by the Joint CUHK-NTEC Clinical Research Ethics Committee (CREC No. 2019.511). For each slide, tiles were extracted into 512×512 resolution at both 5.0 and 20.0 MR as aforementioned. External validation was conducted for tissue classification and KRAS identification. KRAS was selected as external validation biomarkers because patients with KRAS mutants impair response to anti-epidermal growth factor receptor (EGFR) antibody therapy[26].

TABLE 2. The Number of Tiles for Different Datasets Used in Subtask 1

Ensemble	Magnification	Training	Validation	Testing
Net0	5.0	11,130	2,493	2,904
	20.0	155,703	33,321	41,797
Net1	5.0	9,982	2,884	2,476
	20.0	142,363	40,520	35,002
Net2	5.0	10,838	1,603	2,141
	20.0	154,815	22,731	30,198
Net3	5.0	8,854	1,889	3,062
	20.0	123,933	26,456	44,440
Net4	5.0	9,764	1,472	2,268
	20.0	140,916	21,301	33,036
Net5	5.0	8,170	812	3,204
	20.0	115,270	11,848	45,999
Net6	5.0	12,460	933	4,176
	20.0	174,814	11,988	56,643

D. EVALUATION METRICS

For sub-tasks 1 and 2, the prediction generated by the model represented the probability belonging to one class. For sub-task 3, the output from the model was the class number directly owing to the last softmax function. The evaluation metrics were derived from true positive (TP), false positive (FP), true negative (TN) and false negative (FN). Three classification metrics, such as sensitivity, specificity and precision were calculated as follows:

- Sensitivity represents the positive prediction out of all TPs, which is also called as True Positive Rate (TPR). $\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$.
- Specificity represents the negative prediction out of all TNs, which equals to $1 - \text{False Positive Rate (FPR)}$. $\text{Specificity} = \text{TN} / (\text{FP} + \text{TN})$.
- Precision stands for the proportion of TP out of all positive predictions. $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$.

The ROC curve was plotted and the AUROC was calculated to assess the qualitative and quantitative performance of the models. For sub-task 3, the confusion matrix was presented for multi-class classification.

REFERENCES

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] C. C. Y. Poon et al., "AI-doscopist: A real-time deep-learning-based algorithm for localising polyps in colonoscopy videos with edge computing devices," *npj Digit. Med.*, vol. 3, no. 1, May 2020, Art. no. 73.
- [3] R. Dienstmann, L. Vermeulen, J. Guinney, S. Kopetz, S. Tejpar, and J. Tabernero, "Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer," *Nature Rev. Cancer*, vol. 17, no. 2, pp. 79–92, Feb. 2017.
- [4] E. Van Cutsem et al., "Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer," *New England J. Med.*, vol. 360, no. 14, pp. 1408–1417, Apr. 2009.
- [5] A. Cruz-Roa et al., "Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent," *Sci. Rep.*, vol. 7, Apr. 2017, Art. no. 46450.
- [6] N. Coudray et al., "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," *Nature Med.*, vol. 24, no. 10, pp. 1559–1567, Sep. 2018.
- [7] S. Kalra et al., "Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence," *npj Digit. Med.*, vol. 3, Mar. 2020, Art. no. 31.
- [8] Y. Fu et al., "Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis," *Nature Cancer*, vol. 1, pp. 800–810, Aug. 2020.
- [9] J. N. Kather et al., "Pan-cancer image-based detection of clinically actionable genetic alterations," *Nature Cancer*, vol. 1, pp. 789–799, Aug. 2020.
- [10] K. Sirinukunwattana et al., "Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning," *Gut*, vol. 70, no. 3, pp. 544–554, 2021.
- [11] D. Bychkov et al., "Deep learning based tissue analysis predicts outcome in colorectal cancer," *Sci. Rep.*, vol. 8, Feb. 2018, Art. no. 3395.
- [12] J. Guinney et al., "The consensus molecular subtypes of colorectal cancer," *Nature Med.*, vol. 21, no. 11, pp. 1350–1356, Nov. 2015.
- [13] B. Vogelstein, D. Lane, and A. J. Levine, "Surfing the p53 network," *Nature*, vol. 408, no. 6810, pp. 307–310, Nov. 2000.
- [14] P. Wong et al., "Prevalence of early onset colorectal cancer in 397 patients with classic Li-Fraumeni syndrome," *Gastroenterology*, vol. 130, no. 1, pp. 73–79, Jan. 2006.
- [15] M. B. Yurgelun et al., "Germline cancer susceptibility gene variants, somatic second hits, and survival outcomes in patients with resected pancreatic cancer," *Genet. Med.*, vol. 21, no. 1, pp. 213–223, 2019.
- [16] A. D. Singhi et al., "Preoperative next-generation sequencing of pancreatic cyst fluid is highly accurate in cyst classification and detection of advanced neoplasia," *Gut*, vol. 67, no. 12, pp. 2131–2141, 2018.
- [17] W. Abida et al., "Genomic correlates of clinical outcome in advanced prostate cancer," *Proc. Nat. Acad. Sci.*, vol. 116, no. 23, pp. 11428–11436, 2019.
- [18] A. Petitjean, M. I. W. Achatz, A. L. Borresen-Dale, P. Hainaut, and M. Olivier, "TP53 mutations in human cancers: Functional selection and impact on cancer prognosis and outcomes," *Oncogene*, vol. 26, no. 15, pp. 2157–2165, Apr. 2007.
- [19] M. Seri et al., "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes," *Nature Genet.*, vol. 26, no. 1, pp. 103–105, Sep. 2000.
- [20] W. H. L. Kao et al., "MYH9 is associated with nondiabetic end-stage renal disease in African Americans," *Nature Genet.*, vol. 40, no. 10, pp. 1185–1192, Oct. 2008.
- [21] M. J. Ellis et al., "Whole-genome analysis informs breast cancer response to aromatase inhibition," *Nature*, vol. 486, no. 7403, pp. 353–360, Jun. 2012.
- [22] F. Gao et al., "DeepCC: A novel deep learning-based framework for cancer molecular subtype classification," *Oncogenesis*, vol. 8, Aug. 2019, Art. no. 44.
- [23] C. Shyr, M. Tarailo-Graovac, M. Gottlieb, J. J. Y. Lee, C. van Karnebeek, and W. W. Wasserman, "FLAGS, frequently mutated genes in public exomes," *BMC Med. Genomic.*, vol. 7, Dec. 2014, Art. no. 64.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Seattle, WA, USA, 2016*, pp. 770–778.
- [25] R. K. Zhang, Y. L. Zheng, C. C. Y. Poon, D. G. Shen, and J. Y. W. Lau, "Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker," *Pattern Recognit.*, vol. 83, pp. 209–219, Nov. 2018.
- [26] N. Normanno, S. Tejpar, F. Morgillo, A. De Luca, E. Van Cutsem, and F. Ciardiello, "Implications for KRAS status and EGFR-targeted therapies in metastatic CRC," *Nature Rev. Clin. Oncol.*, vol. 6, no. 9, pp. 519–527, Sep. 2009.
- [27] N. N. Mahmoud, S. K. Boolbol, R. T. Bilinski, C. Martucci, A. Chadburn, and M. M. Bertagnolli, "Apc gene mutation is associated with a dominant-negative effect upon intestinal cell migration," *Cancer Res.*, vol. 57, no. 22, pp. 5045–5050, Nov. 1997.
- [28] K. Aoki and M. M. Taketo, "Adenomatous polyposis coli (APC): A multi-functional tumor suppressor gene," *J. Cell Sci.*, vol. 120, no. 19, pp. 3327–3335, Oct. 2007.
- [29] M. K. VanKlompberg, C. O. Bedalov, K. F. Soto, and J. R. Prospero, "APC selectively mediates response to chemotherapeutic agents in breast cancer," *BMC Cancer*, vol. 15, Jun. 2015, Art. no. 457.
- [30] S. Guillemette et al., "Resistance to therapy in BRCA2 mutant cells due to loss of the nucleosome remodeling factor CHD4," *Genes Develop.*, vol. 29, no. 5, pp. 489–494, Mar. 2015.

- [31] L. D. McKenzie et al., "CHD4 regulates the DNA damage response and RAD51 expression in glioblastoma," *Sci. Rep.*, vol. 9, Mar. 2019, Art. no. 4444.
- [32] N. Xu et al., "CHD4 mediates proliferation and migration of non-small cell lung cancer via the RhoA/ROCK pathway by regulating PHF5A," *BMC Cancer*, vol. 20, Mar. 2020, Art. no. 262.
- [33] A. Lievre et al., "KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer," *Cancer Res.*, vol. 66, no. 8, pp. 3992–3995, Apr. 2006.
- [34] E. Valtorta et al., "KRAS gene amplification in colorectal cancer and impact on response to EGFR-targeted therapy," *Int. J. Cancer*, vol. 133, no. 5, pp. 1259–1265, Sep. 2013.
- [35] W. Pao and N. Girard, "New driver mutations in non-small-cell lung cancer," *Lancet Oncol.*, vol. 12, no. 2, pp. 175–180, Feb. 2011.
- [36] M. Laplante and D. M. Sabatini, "mTOR signaling in growth control and disease," *Cell*, vol. 149, no. 2, pp. 274–293, Apr. 2012.
- [37] T. Tian, X. Y. Li, and J. H. Zhang, "mTOR signaling in cancer and mTOR inhibitors in solid tumor targeting therapy," *Int. J. Mol. Sci.*, vol. 20, no. 3, Feb. 2019, Art. no. 755.
- [38] Q. Sheng et al., "An activated ErbB3/NRG1 autocrine loop supports in vivo proliferation in ovarian cancer cells," *Cancer Cell*, vol. 17, no. 3, pp. 298–310, Mar. 2010.
- [39] Y. Samuels et al., "Mutant PIK3CA promotes cell growth and invasion of human cancer cells," *Cancer Cell*, vol. 7, no. 6, pp. 561–573, Jun. 2005.
- [40] I. G. Campbell et al., "Mutation of the PIK3CA gene in ovarian and breast cancer," *Cancer Res.*, vol. 64, no. 21, pp. 7678–7681, Nov. 2004.
- [41] O. Kawano et al., "PIK3CA mutation status in Japanese lung cancer patients," *Lung Cancer*, vol. 54, no. 2, pp. 209–215, Nov. 2006.
- [42] R. Roskoski, "ROS1 protein-tyrosine kinase inhibitors in the treatment of ROS1 fusion protein-driven non-small cell lung cancers," *Pharmacol. Res.*, vol. 121, pp. 202–212, Jul. 2017.
- [43] A. Drilon, C. Jenkins, S. Iyer, A. Schoenfeld, C. Keddy, and M. A. Davare, "ROS1-dependent cancers—Biology, diagnostics and therapeutics," *Nature Rev. Clin. Oncol.*, vol. 18, no. 1, pp. 35–55, Jan. 2021.
- [44] M. Evans et al., "The clinicopathological and molecular associations of PD-L1 expression in non-small cell lung cancer: Analysis of a series of 10,005 cases tested with the 22C3 assay," *Pathol. Oncol. Res.*, vol. 26, no. 1, pp. 79–89, Jan. 2020.
- [45] R. C. Pestana, S. Sen, B. P. Hobbs, and D. V. S. Hong, "Histology-agnostic drug development - considering issues beyond the tissue," *Nature Rev. Clin. Oncol.*, vol. 17, no. 9, pp. 555–568, Sep. 2020.
- [46] R. Roskoski, "Properties of FDA-approved small molecule protein kinase inhibitors: A 2020 update," *Pharmacol. Res.*, vol. 152, Feb. 2020, Art. no. 104609.
- [47] N. Duma, R. Santana-Davila, and J. R. Molina, "Non-small cell lung cancer: Epidemiology, screening, diagnosis, and treatment," *Mayo Clinic Proc.*, vol. 94, no. 8, pp. 1623–1640, Aug. 2019.
- [48] C. E. McCoach et al., "Resistance mechanisms to targeted therapies in ROS1⁺ and ALK⁺ non-small cell lung cancer," *Clin. Cancer Res.*, vol. 24, no. 14, pp. 3334–3347, Jul. 2018.
- [49] E. Cocco, M. Scaltriti, and A. Drilon, "NTRK fusion-positive cancers and TRK inhibitor therapy," *Nature Rev. Clin. Oncol.*, vol. 15, no. 12, pp. 731–747, Dec. 2018.
- [50] Q. X. Hu et al., "ZFHX3 is indispensable for ER beta to inhibit cell proliferation via MYC downregulation in prostate cancer cells," *Oncogenesis*, vol. 8, no. 4, Apr. 2019, Art. no. 28.