Comparison of deep learning architectures for colon cancer mutation detection

Robin Heckenauer
IRIMAS
University of Haute-Alsace
Mulhouse, France
robin.heckenauer@uha.fr

Jonathan Weber
IRIMAS
University of Haute-Alsace
Mulhouse, France
jonathan.weber@uha.fr

Cédric Wemmert
ICube
University of Strasbourg
Strasbourg, France
wemmert@unistra.fr

Caroline Truntzer
Platform of Transfer in Biological Oncology
Georges-François Leclerc Center
Dijon, France
ctruntzer@cgfl.fr

Valentin Derangère
Platform of Transfer in Biological Oncology
Georges-François Leclerc Center
Dijon, France
vderangere@cgfl.fr

François Ghiringhelli
Platform of Transfer in Biological Oncology
Georges-François Leclerc Center
Dijon, France
fghiringhelli@cgfl.fr

Michel Hassenforder
IRIMAS
University of Haute-Alsace
Mulhouse, France
michel.hassenforder@uha.fr

Pierre-Alain Muller
IRIMAS
University of Haute-Alsace
Mulhouse, France
pierre-alain.muller@uha.fr

Germain Forestier
IRIMAS
University of Haute-Alsace
Mulhouse, France
germain.forestier@uha.fr

***Abstract*—Colorectal cancer is responsible of the death of hundred of thousands of people worldwide each year. The histopathological features of the tumor are generally identified from the analysis of tissue taken from a biopsy providing information for selecting the adequate treatment. With the advent of digital pathology, slides of tissue are increasingly available as Whole Slide Images (WSI) allowing their analysis using artificial intelligence algorithms. Additionally, genetic sequencing of the tumor can also be performed to identify specific mutations (e.g. KRAS) and their subtypes (e.g. G12C). While sequencing can take time and is costly, it provides key information on the specific type of cancer. In this paper, we target the identification of key genes and variants using only stained histopathology images using deep neural networks. Predicting gene information only from histopathology images could be easily used routinely and would allow saving the time and cost of tumor sequencing. Experiments performed on 45 colorectal cancer WSI from 45 patients revealed that deep models can correctly classify 90% of patients with the KRAS G12C mutation. Furthermore, we also used CAM and LIME methods to explore the interpretability of the results highlighting which parts of the images were used by the models to predict the presence of specific mutations.**

***Index Terms*—Digital pathology, histopathological images, colorectal cancer, mutation classification, deep learning, explainability**

I. Introduction

The emergence of slide scanners in the late 90's made it possible to digitize glass slides into a very high resolution image named Whole Slide Images (WSI). Examples of patches from such images are shown in Figure 1. This enables to take advantage of artificial intelligence approaches to help pathologists on different tasks like diagnosis.

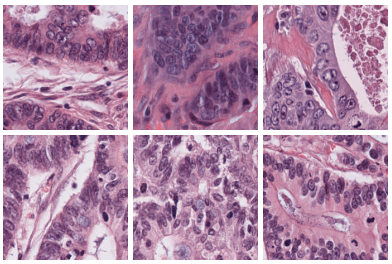


Fig. 1: Examples of patches of tumor areas in colon cancer stained in HES from patient with mutation KRAS.

In its annual report [1], the American Cancer Society (ACS) estimates that by 2022, there will be ∼1,900,000 new cancer cases and ∼600,000 cancer-related deaths in the United States. Colorectal cancer alone is responsible for ∼100,000 new cases and ∼50,000 deaths. Nevertheless, not all colorectal cancers are the same. Indeed, there are different types of cancer and different stages depending on the progress of the disease. Knowing the type of cancer is essential to personalize patients treatment and thus increase chances of survival. In the context of a microscopic analysis, a pathologist is able to determine a lot of information about the cancer and its progress from a glass slide. Deep learning methods are also able to perform such tasks [2]. However, visual tissue analysis is not sufficient

DOI: https://doi.org/10.1109/CBMS58004.2023.00244

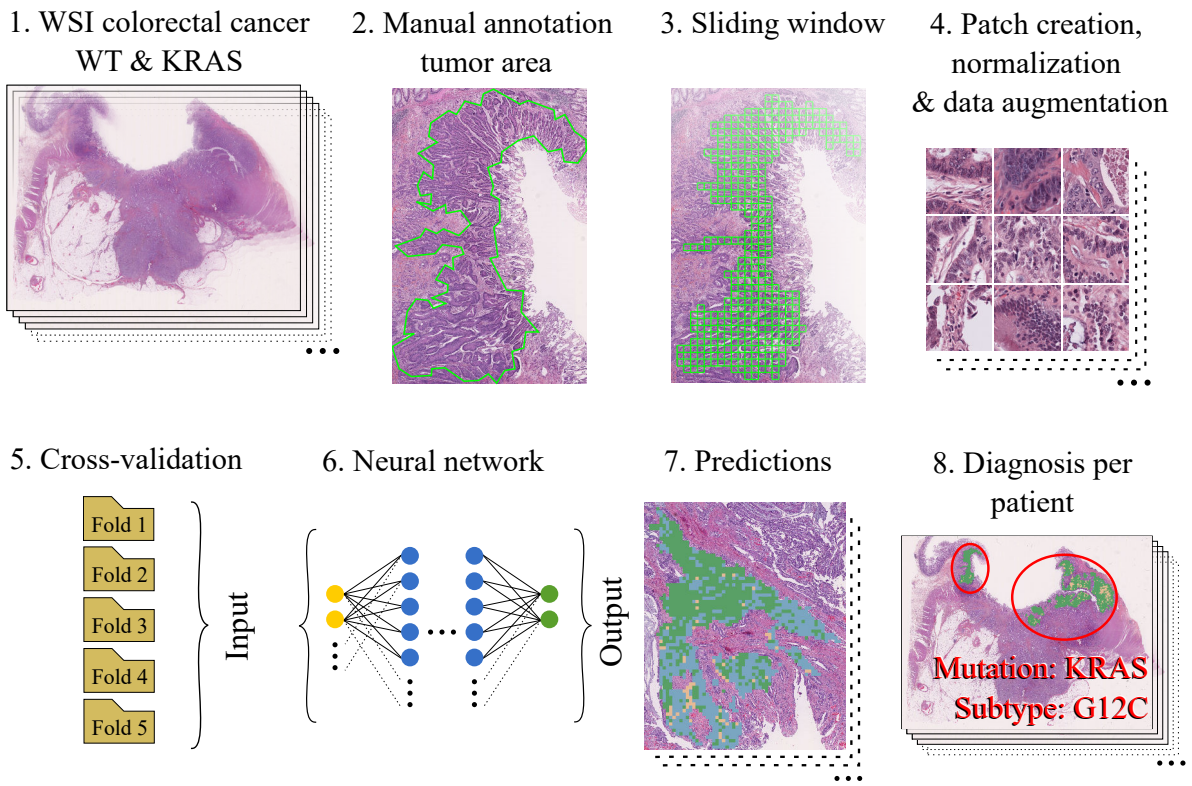


Fig. 2: Our workflow. A collection of colorectal cancer WSI containing KRAS mutation and WT (1) is tumor annotated (2). We use a sliding window (3) to create patches size 224∗224 (4). These patches are organized in five folds (5) to perform cross-validation. Then, we train a deep neural network (6) to predict mutation and variant (7). Finally, we use the predictions to determine the patient's mutation (8).

to know which genes are mutated or wild-type (WT i.e. non-mutated) and which variants are present. This step is crucial to determine certain types of cancer and guide patients to the best treatment. In the case of colorectal cancer, 3% of cancers have KRAS G12C (Glycine 12 to Cysteine) mutation. This rare mutation plays a key role in cell proliferation and is responsible for an extremely aggressive forms of cancer. However, there are treatments that can improve patients' chances of survival and that can be prescribed from the moment the mutation is identified [3]. Genome information can be obtained by a DNA sequencer, but remains expensive and time consuming

In 2019, N. Coudray et al. [4] were the first to successfully apply a deep learning method to predict mutations directly in lung WSI. In the wake of this work, many deep learning-based approaches have been proposed to predict genetic information [4]–[9]. Among them, some approaches focus on mutations in colorectal cancer [6], [8]. These encouraging results are surprising because pathologists are not able to determine visually the presence or absence of a mutation in a gene from a microscopic analysis of a tissue. Yet, deep neural networks seem to be able to do it. Therefore, deep learning methods have the potential to determine genetic information without going through time-consuming and expensive DNA sequencing. Moreover, networks could also allow us to understand the biological structures associated with mutations through the field of explainability. Despite initiatives to understand network predictions using explainability methods [6], [7], the structures that lead to the prediction of mutated genes and variants are still poorly understood [7].

In this paper, we propose:

- To compare different deep learning architectures on the task of classification of KRAS mutation in histopathological images of colorectal cancer.
- To classify the G12C variant of the KRAS gene given its interest in the selection of a treatment that increases the chances of survival.
- To use the CAM [10] and LIME [11] explainability methods to visualize the areas of patches that motivate a network's decision making.

The paper is organized as follows. First, we present related work on deep learning approaches to predict mutations in section II. Then, we detail our proposed approach and our colon cancer WSI dataset in section III, before comparing different architectures in section IV.

To ensure reproducibility, we also release the source code of our method here [1].

[1]https://github.com/RobinHCK/Comparison-of-deep-learning-architectures-for-colon-cancer-mutation-detection

II. Related works

In the last years, approaches based on deep neural networks have been investigated to predict genetic information from histopathological images. However, there is still a lack of understanding of which features the networks rely on to perform their predictions. To cope with this issue, the field of explainability explore techniques allowing to better understand how the detection and decision are made by the deep models.

In 2019, [4] was the first to show the potential of deep neural networks on the task of classifying mutated genes in histopathological tissue patches. In the wake of this work, many deep learning-based approaches have been proposed to predict genetic information [4]–[8], [12]–[19]. These works mainly focus on organs with the most deadliest cancers [20] such as lung [4], [7], [17], colon [6], [8], liver [5], [13], breast [14], [15], pancreas [16], prostate [12], uterus [18], kidney [19]. Starting from these organs, there are two types of methods to collect medical images. On the one hand, invasive methods allow to obtain histopathological images as in [4]–[6], [12]–[15], [18], [19]. On the other hand, there are also non-invasive methods that allow to create computed tomography (CT) images as in [7], [8], [16], [17]. In both cases, medical images are used in a similar way to train deep neural networks.

Among the proposed approaches, we mainly find methods based on ResNet [8], [12], [14]–[19], which use skip connections to reduce the vanishing gradient phenomenon. Then, the Inception-v3 architecture, which was notably used by [4] to show the feasibility of predicting mutations in histopathological images, was also taken up [5], [18]. The specificity of Inception architectures is to use filters of different sizes for each level to reduce overfitting. In addition to these works, mainly based on ResNet and Inception, there are articles that use other architectures: CNN [13], [21], DNN [22], DenseNet [7], EfficientNet [17]. Note that these deep neural networks are mainly used to directly classify patches according to their mutations and subtypes. However, some works try to include patient data in addition to patches in order to give more context to the network, as in [17]. Given the current work, the many layers of deep learning architectures seem to be able to extract abstract features specific to mutations, however to our knowledge, there is no comparison between these architectures on the mutation prediction task.

Although the results of the networks on the genetic information prediction task are promising, the existing methods struggle to justify the predictions of deep neural networks, as described in [5]. In order to bypass this black box effect, [6] studies the relationships between some genes using t-SNE [23]. In parallel, [7], [22] explored the use of CAM [10] to try to explain visually the predictions of the networks. Indeed, explainability methods such as CAM and LIME [11] allow to visualize the impact of each part of a patch in its final classification result. On the one hand, it is possible to use these results to improve the learning of a network as shown in [24]. On the other hand, these methods are particularly useful for understanding the decisions of networks that tend to become more complex to handle more difficult tasks [25], such as mutation classification. Despite these initiatives, the behavior of networks on the mutation prediction task is not well understood, in particular because of the abstract nature of the features due to the large number of parameters in the networks, as described in [7].

TABLE I: Folds per split. Distribution of folds in training, validation and test splits for each network in the cross-validation.

	#Folds in		
#Network	Test	Validation	Train
1	1	2	3, 4, 5
2	2	3	4, 5, 1
3	3	4	5, 1, 2
4	4	5	1, 2, 3
5	5	1	2, 3, 4

III. Materials and methods

A. Materials

The AiCOLO dataset is a private dataset which contains 143 colorectal cancer WSI from 105 patients. After removing the WSI with defects (blurring, staining, etc), we keep 45 WSI from 45 patients. All the WSI were stained with Hematoxylin Eosin Saffron (HES), before being scanned using a Hamamatsu photonics scanner at X20 magnification and a spatial resolution of 0.454μm/pixel. The WSI come from the Georges-François Leclerc Center. The AiCOLO dataset also contains annotations of all tumor areas as well as genetic information obtained by DNA sequencing, such as the mutated genes (WT, BRAF, KRAS, NRAS, etc). For each WSI containing KRAS gene, we also have a specific variant (G12C, G12D, G12V, etc).

We use a sliding window with 0 overlap as shown in Figure 2 (3) to create patches of size 224×244 pixels at X20 magnification. As a result, we label all patches of a WSI with the same mutation and the same variant. To reduce bias in our data, we apply the Macenko [26] normalization method as well as other on-the-fly data augmentation methods during the training step, such as flip, rotation and shear.

The AiCOLO dataset contains a majority of WSI with the KRAS mutation. Considering that in this study we focus on the KRAS G12C mutation, we create a first dataset of 76,000 patches for the mutation classification task containing 38,000 KRAS patches (KRAS) and 38,000 non KRAS patches (WT). Then, considering the low amount of data of the other variants, we choose to group the minority classes into one class, named "Others". Thus, to classify the variants, a second dataset of 70,500 patches is created containing 23,500 KRAS G12C patches (G12C), 23,500 other variant patches (Others) and 23,500 patches without mutation (WT). The patches of the G12C, Others and WT classes were created from 45 WSI (13 WSI G12C, 13 WSI Others and 19 WSI WT). During the process of patches creation, we randomly and proportionally remove patches in each WSI to balance the data.

TABLE II: Comparison of architectures for KRAS mutation classification. Comparison of the accuracy of the architectures on the task of classification of patches with KRAS mutation and WT.

Architectures	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	All Folds
InceptionResNetV2	0.782	0.745	0.784	0.681	0.850	**0.769** *(±0.056)*
MobileNetV2	0.832	0.667	0.816	0.648	**0.866**	0.766 *(±0.090)*
DenseNet201	0.783	0.650	**0.845**	0.681	0.833	0.758 *(±0.079)*
ResNet152V2	0.780	0.670	0.830	0.646	0.853	0.756 *(±0.084)*
NASNetLarge	**0.861**	0.571	0.811	0.670	0.845	0.752 *(±0.113)*
Inception-v3	0.788	**0.763**	0.528	**0.710**	0.809	0.720 *(±0.101)*
EfficientNetV2L	0.764	0.496	0.709	0.626	0.500	0.619 *(±0.108)*
VGG19	0.500	0.500	0.500	0.500	0.500	0.500 *(±0)*

TABLE III: Comparison of architectures for KRAS G12C variant classification task. We compare the accuracy of the architectures on classification tasks of the G12C variant of the KRAS gene, per patch and per WSI. The prediction per WSI (i.e. per patient) is computed from the classification results per patch.

	classification of patches	classification of WSI				
	G12C vs		NotG12C		G12C vs	G12C vs
Architectures	Others vs WT	G12C	Others	WT	Others vs WT	NotG12C
MobileNetV2	**0.574** *(±0.135)*	**0.923**	0.462	0.462	0.615	**0.897**
InceptionResNetV2	0.538 *(±0.113)*	**0.923**	**0.538**	0.**538**	**0.667**	0.821
NASNetLarge	0.483 *(±0.111)*	0.846	0.385	0.538	0.590	0.769
ResNet152V2	0.478 *(±0.151)*	0.769	0.231	0.462	0.487	0.746
Inception-v3	0.482 *(±0.079)*	0.692	0.154	0.692	0.513	0.641
DenseNet201	0.428 *(±0.166)*	0.615	0.231	0.385	0.410	0.641
EfficientNetV2L	0.463 *(±0.115)*	0.846	0.462	0.000	0.536	0.615
VGG19	0.333 *(±0)*	0.154	0.385	0.385	0.308	0.590

B. Methods

For all our experiments, we use cross-validation with five folds (three training folds, one validation fold, and one test fold) as shown in Table I to take full advantage of our data and reduce overfitting. We distribute the WSI of the different classes equally among the five folds to avoid our networks associating the most represented class to all the data. Moreover, we take care to avoid data leakage by putting the patches from a WSI (i.e. patient) in a single fold so that it is impossible for a network to test patches from a patient that it could have seen during the training step. Each fold is used once to test a model trained on all the other folds. Then, the results of the five models on the five folds are aggregated.

As we want to detect the G12C subtype of the KRAS mutation, we proceed in two steps. First, we train our model to discriminate the KRAS mutation (KRAS vs WT dataset). Then we repeat our experiments to discriminate the G12C subtype of the KRAS mutation (G12C vs Others vs WT dataset).

In both cases, we compute the accuracy over all folds from the patches that are well and poorly classified by the network. Then, we assign to each WSI (i.e. patient) the most predicted class among the patches that compose it. Again, we compute the accuracy on all the WSI well and poorly classified by the network. First, we detail the accuracy obtained for each class (G12C, Others and WT), then we present the accuracy for the tasks G12C vs Others vs WT and G12C vs NotG12C.

IV. Experiments & Discussion

A. Classification of the KRAS mutation

We choose to compare the performance of deep learning architectures that we selected based on the literature: DenseNet201 [27], EfficientNetV2L [28], Inception-v3 [29], InceptionResNetV2 [30], MobileNetV2 [31], NASNetLarge- [32], ResNet152V2 [33], and VGG19 [34]. The following hyper-parameters were obtained using random search optimization : Learning rate 0.01; Epochs 30; Batch size 32; Cost function categorical crossentropy; Optimization function RMSprop; Dropout rate 0.1; L2 Regularization 0.0001.

We report the results of our networks on the KRAS vs WT patch classification task in Table II. Then, after finding that the networks were able to differentiate the KRAS mutation from the wild type, we repeated our experiments on the G12C subtype classification task (G12C vs Others vs WT).

B. Classification of the KRAS G12C subtype

The accuracy of the different architectures are reported in Table III. First, the accuracy of the networks on the G12C vs Others vs WT patch classification task are between 46.3% and 57.4%. Only VGG19 fails to learn. Secondly, when we look at the WSI classification results, still on the G12C vs Others vs WT task, we see that the accuracy are higher than on the patch classification. Third, if we look in detail at the accuracy for each of the classes of the WSI classification (as presented in the G12C, Others and WT columns), we see that the classes containing the other variants (Others) and the wild-type (WT) are confused by the networks. We believe this is

due to the limited data we have for these classes. Since we focus on the detection of the KRAS G12C mutation because of the aggressive cancer forms it creates, we can reduce our problem to only two classes by combining the Others and WT classes. Thus, in the G12C vs NotG12C column, we can see that our networks correctly recognize the G12C subtype.

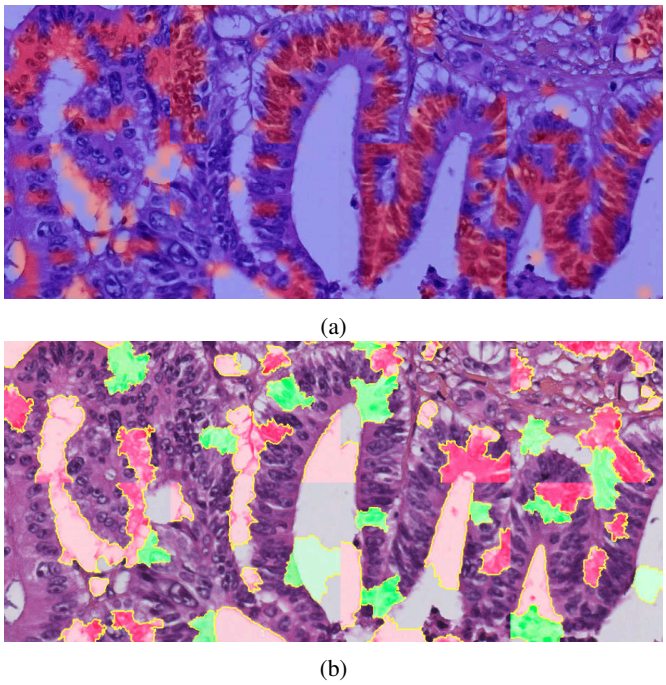
Several deep neural networks achieve high accuracy on KRAS G12C mutation classification task per WSI (G12C vs NotG12C) with 74.6% for ResNet152V2, 76.9% for NASNetLarge, 82.1% for InceptionResNetV2 and 89.7% for MobileNetV2. It appears that the MobileNet specific convolutions and the residual layers introduced with ResNet, allow the extraction of relevant mutation related features. We note that the accuracy with which our networks classify the KRAS G12C mutation is comparable with the results reported by [6] on colorectal cancer WSI HES. To the best of our knowledge, we propose the best results in the literature on the KRAS G12C mutation detection task per patient using our MobileNetV2 models.

In this study, we successfully predicted the G12C subtype of the KRAS gene. However, this was not the case for the other available variants. It is not clear whether all mutations could be detected on WSI, or if it is due to a limitation of our experimentation conditions. The difficulties encountered by our networks can be explained by the poorly labeled training data that we give to the networks. Indeed, we assign to all patches of a WSI the same class, but the signatures of mutations and variants are not necessarily present in each patch. The lack of data can also be at the origin of the failure to classify the least represented mutations and variants. For similar reasons, we believe that our networks perform better on patient-based classification than on patch-based classification because it smoothes out network errors. Although it is possible that not all patches show a signature of the gene we are looking for, a majority of patches with a magnification level of X20 are likely to contain traces of a mutation because they still allow the networks to learn about the characteristics of the mutation. This kind of approach could be used in laboratory routines to quickly and at low cost obtain genetic information on a glass slide, as is the case for other medical image analysis methods described by [35]. Nevertheless, there are still technical obstacles, for example, we have found that traces of mucins on tissues are often interpreted as a variant from the class Others by the networks.

In our experiments, we focus on KRAS G12C for its clinical interest but also because our dataset contains enough diversity for learning purposes. Nevertheless our subsidiary experiments suggest that using more genes and variants when training a network can improve network performance.

C. Explainability with CAM and LIME

Finally, we applied CAM and LIME explainability methods to try to understand which visual features helped the network to make its decision (see Figure 3). We observe that CAM highlights areas that are associated with tumors. However, after analysis of these results by a pathologist, we could not visually determine the structures that discriminate the KRAS mutation or the G12C variant. We believe that the difficulty in interpreting the results given by the explainability methods comes from the biases of our networks. It seems that the proliferation of cells with a mutated gene produces patterns that allow AIs to predict mutations. However, these patterns are too complex or too small to be yet identified visually by a pathologist with a microscope. Thus, it is relevant to use deep learning methods to detect what an expert is not able to see.



(a)

(b)

Fig. 3: Example of CAM (a) and LIME (b) results on network predictions. Highlighted areas show where the network focuses to classify KRAS G12C.

V. Conclusion

In recent years, the advent of deep learning methods has made it possible to tackle more complex tasks that force networks to learn increasingly abstract features. In this context, the prediction of genetic information shows encouraging results, however, understanding the relationships between network predictions and biological structures remains complex. We proposed a method to predict KRAS mutation and G12C subtype in colorectal cancer with neural networks. Then, we attempt to identify the structures that the networks associate with the KRAS G12C mutation using the CAM and LIME explainability methods. We believe that our method is transferable to other organs and can help to predict genes and variants. Future work will focus on adding patient data as input to the networks to add more context, before investigating the relationships between biological structures and genetic mutations using methods to understand network features.

ACKNOWLEDGMENTS

This work was supported by the AiCOLO project funded by INSERM/Plan Cancer. The computational resources were provided by the Mesocentre of the University of Strasbourg.

REFERENCES

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," *CA: A Cancer Journal for Clinicians*, vol. 72, no. 1, pp. 7–33, 2022.

[2] A. B. Hamida, M. Devanne, J. Weber, C. Truntzer, V. Derangère, F. Ghiringhelli, G. Forestier, and C. Wemmert, "Deep learning for colon cancer histopathological images analysis," *Computers in Biology and Medicine*, vol. 136, p. 104730, 2021.

[3] M. Nagasaka, Y. Li, A. Sukari, S.-H. I. Ou, M. N. Al-Hallak, and A. S. Azmi, "Kras g12c game of thrones, which direct kras inhibitor will claim the iron throne?," *Cancer treatment reviews*, vol. 84, p. 101974, 2020.

[4] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature medicine*, vol. 24, no. 10, pp. 1559–1567, 2018.

[5] M. Chen, B. Zhang, W. Topatana, J. Cao, H. Zhu, S. Juengpanich, Q. Mao, H. Yu, and X. Cai, "Classification and mutation prediction based on histopathology h&e images in liver cancer using deep learning," *NPJ precision oncology*, vol. 4, no. 1, pp. 1–7, 2020.

[6] C. Bian, Y. Wang, Z. Lu, Y. An, H. Wang, L. Kong, Y. Du, and J. Tian, "Immunoaizer: A deep learning-based computational framework to characterize cell distribution and gene mutation in tumor microenvironment," *Cancers*, vol. 13, no. 7, p. 1659, 2021.

[7] S. Wang, J. Shi, Z. Ye, D. Dong, D. Yu, M. Zhou, Y. Liu, O. Gevaert, K. Wang, Y. Zhu, *et al.*, "Predicting egfr mutation status in lung adenocarcinoma on computed tomography image using deep learning," *European Respiratory Journal*, vol. 53, no. 3, 2019.

[8] K. He, X. Liu, M. Li, X. Li, H. Yang, and H. Zhang, "Noninvasive kras mutation estimation in colorectal cancer using a deep learning method based on ct imaging," *BMC medical imaging*, vol. 20, no. 1, pp. 1–9, 2020.

[9] M. Liu, S. Wang, H. Yu, Y. Zhu, L. Wang, M. Zhang, Z. Wu, X. Li, W. Li, and J. Tian, "A lung-parenchyma-contrast hybrid network for egfr gene mutation prediction in lung cancer," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, IEEE, 2022.

[10] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Is object localization for free?-weakly-supervised learning with convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 685–694, 2015.

[11] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you? explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.

[12] A. J. Schaumberg, M. A. Rubin, and T. J. Fuchs, "H&e-stained whole slide image deep learning predicts spop mutation state in prostate cancer," *BioRxiv*, p. 064279, 2018.

[13] H. Liao, Y. Long, R. Han, W. Wang, L. Xu, M. Liao, Z. Zhang, Z. Wu, X. Shang, X. Li, *et al.*, "Deep learning-based classification and mutation prediction from histopathological images of hepatocellular carcinoma," *Clinical and translational medicine*, vol. 10, no. 2, 2020.

[14] H. Qu, M. Zhou, Z. Yan, H. Wang, V. K. Rustgi, S. Zhang, O. Gevaert, and D. N. Metaxas, "Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning," *NPJ precision oncology*, vol. 5, no. 1, pp. 1–11, 2021.

[15] X. Wang, C. Zou, Y. Zhang, X. Li, C. Wang, F. Ke, J. Chen, W. Wang, D. Wang, X. Xu, *et al.*, "Prediction of brca gene mutation in breast cancer based on deep learning and histopathology images," *Frontiers in Genetics*, p. 1147, 2021.

[16] X. Chen, X. Lin, Q. Shen, and X. Qian, "Combined spiral transformation and model-driven multi-modal deep learning scheme for automatic prediction of tp53 mutation in pancreatic cancer," *IEEE Transactions on Medical Imaging*, vol. 40, no. 2, pp. 735–747, 2020.

[17] S. Tripathi, A. I. Augustin, E. J. Moyer, A. Zavalny, S. Dheer, R. Sukumaran, D. Schwartz, B. Gorski, F. Dako, and E. Kim, "Radgennets: Deep learning-based radiogenomics model for gene mutation prediction in lung cancer," *bioRxiv*, 2022.

[18] R. Hong, W. Liu, D. DeLair, N. Razavian, and D. Fenyö, "Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models," *Cell Reports Medicine*, vol. 2, no. 9, p. 100400, 2021.

[19] S. Tabibu, P. Vinod, and C. Jawahar, "Pan-renal cell carcinoma classification and survival prediction from histopathology images using deep learning," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.

[20] J. Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray, "Cancer statistics for the year 2020: An overview," *International journal of cancer*, vol. 149, no. 4, pp. 778–789, 2021.

[21] K. Lee, B. Kim, Y. Choi, S. Kim, W. Shin, S. Lee, S. Park, S. Kim, A. C. Tan, and J. Kang, "Deep learning of mutation-gene-drug relations from the literature," *BMC bioinformatics*, vol. 19, no. 1, pp. 1–13, 2018.

[22] R. R. van de Leur, K. Taha, M. N. Bos, J. F. van der Heijden, D. Gupta, M. J. Cramer, R. J. Hassink, P. van der Harst, P. A. Doevendans, F. W. Asselbergs, *et al.*, "Discovering and visualizing disease-specific electrocardiogram features using deep learning: proof-of-concept in phospholamban gene mutation carriers," *Circulation: Arrhythmia and Electrophysiology*, vol. 14, no. 2, p. e009056, 2021.

[23] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.

[24] M. A. Gulum, C. M. Trombley, and M. Kantardzic, "Improved deep learning explanations for prostate lesion classification through grad-cam and saliency map fusion," in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 498–502, IEEE, 2021.

[25] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," *Journal of Artificial Intelligence Research*, vol. 70, pp. 245–317, 2021.

[26] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *2009 IEEE international symposium on biomedical imaging: from nano to macro*, pp. 1107–1110, IEEE, 2009.

[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[28] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International Conference on Machine Learning*, pp. 10096–10106, PMLR, 2021.

[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

[30] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[31] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

[32] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697–8710, 2018.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, pp. 630–645, Springer, 2016.

[34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[35] S. Panda, M. Jangid, and A. Jain, "A comprehensive review on the significance and impact of deep learning in medical image analysis," in *2021 International Conference on Technological Advancements and Innovations (ICTAI)*, pp. 358–366, IEEE, 2021.