

Identifying Cancer Subtypes through FGFR3 Mutation in Bladder Cancer using Histopathological Images

By: Mothishwara Navinkumar Natarajan
Supervisors: Dr. Andreas Karwath & Dr. Roland Arnold

Table of Contents

1. Abstract:	2
2. Introduction:.....	2
3. Materials and Methods:.....	4
3.1 Dataset:	4
3.2 Workflow:	5
3.3 Foundation Level:.....	6
3.3.1 Data Pre-processing:	6
3.3.2 Feature extraction/vector embedding:.....	6
3.3.3 Slide level vector extraction:.....	7
3.4 Task Level:.....	9
3.4.1 Feature visualization:	9
3.4.2 Model Building:	9
3.4.2 Interpretation:.....	10
4. Results:.....	10
4.1 Top 25 Informative Tiles:	10
4.2 Feature Visualization:	10
4.3 Predictive Modelling:	12
4.4 Interpretive visualization:.....	12
5. Discussion:	15
6. Conclusion:	17
7. References:.....	17

1. Abstract:

Bladder cancer is the 10th most common form of cancer, accounting for 550,000 newly diagnosed cases and 220,000 deaths globally. Since Bladder cancer is highly heterogeneous, it is important to identify the subtype of the cancer for targeted treatment strategies that lead to better outcomes. Based on the invasive nature of the tumor, Bladder cancer can be subtyped into Non-Muscle Invasive Bladder Cancer (NMIBC) and Muscle Invasive Bladder Cancer (MIBC). With advances in sequencing technologies, several studies have reported that the Fibroblast Growth Factor Receptor 3 (FGFR3) mutation is highly frequent in NMIBC and very rarely seen in MIBC. Therefore, identifying the mutation earlier can lead to faster determination of the cancer subtype, which helps in treatment directions. However, in reality, the identification of the FGFR3 mutation is highly reliant on expression studies, such as RNA sequencing, which require a significant amount of time and effort. In this study, we examined whether it is possible to identify the FGFR3 mutation in routinely collected Histopathological Whole Slide Images (WSI) of bladder cancer using deep learning. To achieve this, we utilized pre-trained foundational models based on the Vision Transformer architecture. These models have been trained on thousands of histopathological images obtained from several cancers to learn general, high-quality feature representations in the form of vectors. By using these vectors as base features, a Multi-Layer Perceptron (MLP) neural network is created to model the relationship between the FGFR3 mutation and the images. The model achieved an average Area under the curve (AUC) of 0.81 (+/-) 0.068 across 10 different trials, indicating that there are visual features in the images that help identify the presence of the FGFR3 mutation. Interpretive visualizations were then created to understand the model and examine intra-cancerous heterogeneity between both mutation-present and wild-type classes, which showed that cancerous regions play a vital role in differentiating between the FGFR3 mutation and wild-type classes.

2. Introduction:

Bladder cancer is one of the most prevalent cancers in the world, being the most common malignancy in women and the fourth most common malignancy in men (Lenis *et al.*, 2020). As of 2018, there were roughly 550,000 new cases reported for bladder cancer, with 220,000 deaths annually. (Tran *et al.*, 2021; Dyrskjot *et al.*, 2023). Bladder cancer is highly heterogeneous with different kinds of subtypes. Based on the local invasion of the tumor, bladder cancer can be divided into two subtypes. Non-Muscle Invasive Bladder Cancer (NMIBC) and Muscle Invasive Bladder Cancer (MIBC) (Noeraparast *et al.*, 2024). Non-muscle invasive bladder is the most common out of the two, accounting for roughly 80% of the cases, and the remaining 20% being Muscle Invasive (Noeraparast *et al.*, 2024). The emergence of next-generation sequencing technologies has led to a better understanding of bladder cancer biology and allowed researchers to establish a molecular taxonomy for bladder cancer. This particular effort showed that bladder cancer is one of the most frequently mutated human cancers (Alexandrov *et al.*, 2013; Lawrence *et al.*, 2013; Tran *et al.*, 2021).

One such example is Fibroblast growth factor receptor 3 (FGFR3). FGFR3 is a central pathway responsible for cell surface receptors, playing an important role in cell proliferation and differentiation. Mutations in the FGFR3 gene seem to be highly frequent in NMIBC, while being very rarely seen in MIBC (Alexandrov *et al.*, 2013; Tran *et al.*, 2021; Dyrskjøl *et al.*, 2023). Despite several advances in the understanding of bladder cancer and focused treatment for the FGFR3 mutation, the 5-year survival rate for NMIBC remains the same (75% to 95%) (Tran *et al.*, 2021). One of the main reasons behind this is the need for early diagnosis and subtyping of bladder cancer for targeted treatment strategies, which can improve survival rates (Bannier *et al.*, 2024).

Therefore, in this study, we try to examine if it is possible to identify the FGFR3 mutation using regularly collected Histopathological images of Bladder cancer, which can then lead to the timely determination of the bladder cancer subtype and help in selecting targeted treatment strategies if the mutation is present.

In recent years, foundational models have been gaining increased interest in the field of computer vision, especially in the field of computational pathology (Ochi, Komura and Ishikawa, 2025). The notion behind foundational models is that instead of training models to learn from the images for a very specific task using strong supervision, the model is trained in such a fashion that the main aim is to learn general, high-quality feature representations capturing contextual and morphological information. This is usually done using a weakly/self-supervised learning approach on very large amounts of data (El Nahhas *et al.*, 2025). In this case, the data is made of histopathological images from multiple cancers. Here, pan-cancer is used because it makes the model robust to different kinds of cancer by forcing it to learn variations in different forms of cancer. Hence, the foundation models are task agnostic (Lu *et al.*, 2024). The feature representation obtained using the foundational models can then be used for training less complex models that are for multiple task-specific problems, thereby creating a foundation. Some recent studies involving foundational models, subtyping, and mutational prediction in cancer are discussed below.

Bannier *et al.*, 2024 utilized a foundational model for computational pathology based on a vision-transformer architecture called H-optimus0 and a U-Net convolution model to predict FGFR3 mutation using tiles of bladder cancer images obtained from The Cancer Genome Atlas (TCGA). The predictions were made at the tile level and then aggregated to produce image level mutation prediction. The model performed with an average AUC of 0.8 across different validation sets (Bannier *et al.*, 2024).

Loeffler *et al.*, 2022 used deep convolutional neural networks on TCGA bladder cancer image tiles to detect FGFR3 mutation. The study also tried to identify features that play a pivotal role in identifying the existence of the FGFR3 mutation. The model performed with an AUC of 0.76 (Loeffler *et al.*, 2022b).

Wang *et al.*, 2025 predicted the molecular consensus subtypes (CMS) and assessed if the digital CMS score correlates with patient survival in HPV positive Cervical Squamous cell carcinoma

using the routine Whole Slide Images obtained from several data sources that include TCGA. This study used a deep learning multi-instance learning model called TripletMIL to perform the classification and found that the performance was comparable with molecular-based CMS classification (Wang *et al.*, 2025).

Sirinukunwattana *et al.*, 2021 developed a deep-learning model that predicts the four consensus molecular subtypes (CMS1 - CMS4) of colorectal cancer directly from routine H&E-stained whole-slide images (WSIs), bypassing the need for costly RNA-based CMS classification, and to resolve unclassifiable or heterogeneous cases. The study used a training strategy called domain adversarial learning by incorporating data from multiple sources, such as FOCUS for training and TCGA and GRAMPIAN for validation. The model produced high performances of AUC 0.84 and 0.85 in TCGA and GRAMPIAN, respectively (Sirinukunwattana *et al.*, 2021).

Chen *et al.*, 2020 employed the deep learning Inception v3 model (Szegedy *et al.*, 2016) to distinguish benign liver tissue vs hepatocellular carcinoma (HCC) and classify HCC into histopathological grades. The study also created a classifier to predict the mutation status of ten prognostically important genes from Histopathology images for liver cancer obtained from Genomic Data Commons (GDC) (Chen *et al.*, 2020).

Most of the studies above used deep convolutional neural networks to predict the mutation/subtype of the image on a smaller tile level and then aggregated them. In our approach, we utilize a combination of CNNs and SwinTransformer (Liu *et al.*, 2021) architecture to extract tile features and a foundation model called CHIEF (Wang *et al.*, 2024) to aggregate tile-level features into a slide-level representation before prediction to capture intra-relationship and capture information on a global scale.

3. Materials and Methods:

3.1 Dataset:

The data for the analysis was acquired privately. The dataset comprises digital histopathological tissue samples of bladder cancer as Whole Slide images (WSI) (Zarella *et al.*, 2018) . The samples were scanned at a magnification level of 0.26 Microns per pixel at an approximate resolution of 145,000x87000. The dataset consists of 300 whole-slide images of bladder cancer tissue, along with the FGFR3 mutation status as labels. Of the 300 Images, 51 percent of the samples had the FGFR3 mutation present while the rest 49 percent did not have the FGFR3 mutation. Each slide has a marked region of interest that represents the cancerous tissue.

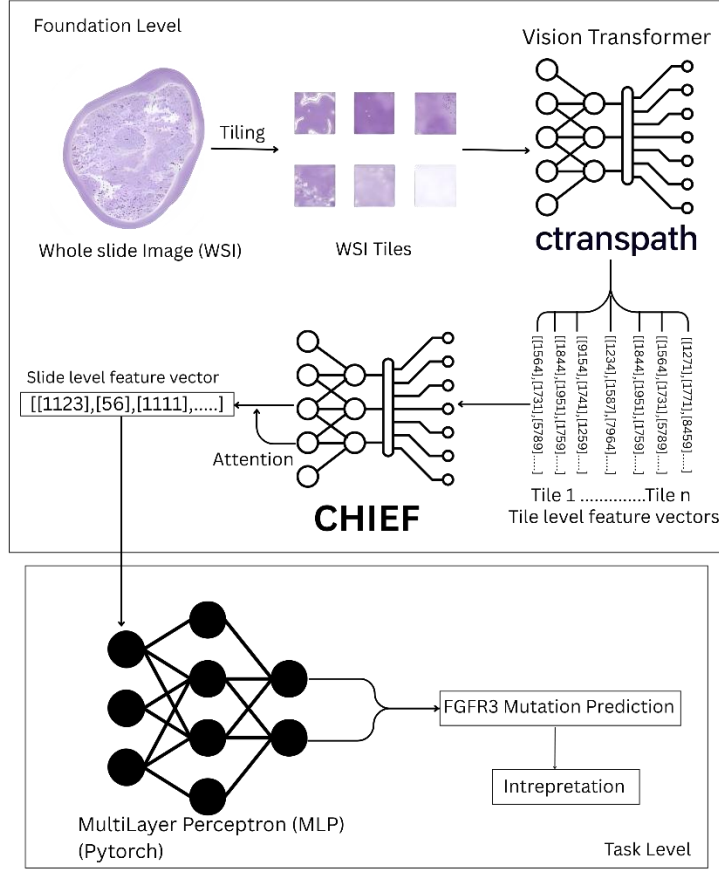


Figure 1: Workflow abstract for predicting FGFR3 mutation in bladder cancer, Whole slide images. Workflow is divided into two stages, the foundational level and the task level. The foundational level aims at extracting general, high-quality feature representations. Task-level aims at predicting the mutation using features obtained from the foundational level.

3.2 Workflow:

The workflow of FGFR3 mutation prediction in bladder cancer images is divided into two levels, the foundation level and the task-specific level, as indicated in Figure 1. The Foundation level mainly focuses on extracting high-quality general-purpose feature representations from raw Whole Slide Images of bladder cancer that act as a base set of features for the downstream task. Therefore, the Foundation level is usually task agnostic. In the task-specific level, the main focus is to implement a model that predicts the FGFR3 mutation status using the features extracted from the foundation level.

3.3 Foundation Level:

3.3.1 Data Pre-processing:

Due to the high resolution of the images, it might not be optimal to use the images directly for the analysis, as they are dominated by background regions, which have no clinical relevance. This necessitates the images to be pre-processed so that background regions can be eliminated, which reduces the time required for processing the slide images (Faghani *et al.*, 2023). Therefore, the slides are preprocessed using two steps. First, each slide image is broken down into several smaller square tiles of size 224x224 pixels at a magnification level of 2 microns per pixel. Which means each tile covers a total width and height of 448 microns ($224\text{px} * 2(\text{microns per pixel})$). The Images are broken down into tiles so that they are suitable to be used with vision transformers, which is the preferred model architecture for this analysis. The transformer architecture was originally designed for natural language processing tasks, where the inputs are a sequence of tokens (words/sub-words) (Vaswani *et al.*, 2017). In order to adapt the transformer architecture to images, the images must be tiled, which can then be flattened to replicate the sequence of tokens (Dosovitskiy *et al.*, 2021). After breaking the WSI images into tiles, the tiles undergo a selection process to eliminate non-informative background regions (Faghani *et al.*, 2023). Any tile greater than the threshold of 95% background region (based on Otsu Thresholding) (White region) was not considered for analysis and was eliminated (Otsu, 1979). Figure 2 shows the original slide and its processed equivalent, where the tiles with over 95 percent region are eliminated.

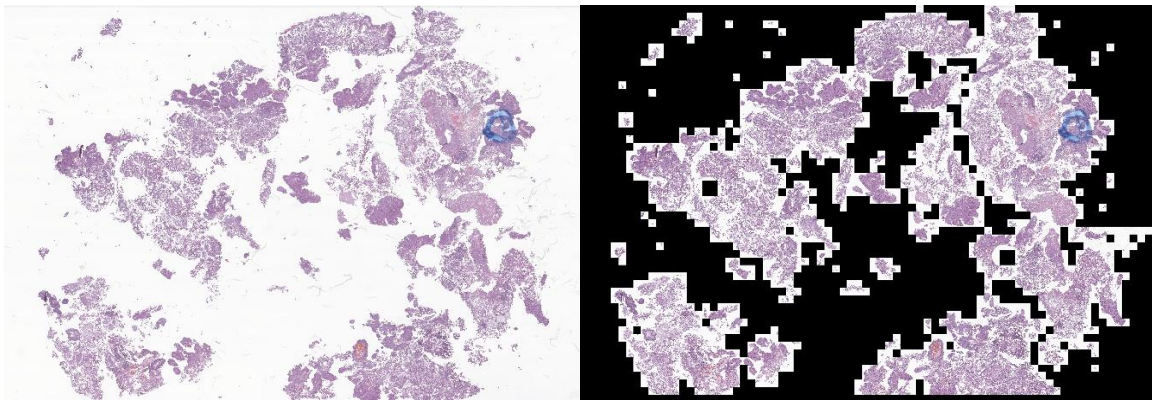


Figure 2: Bladder cancer whole slide image before (left) and after processing (right). Any tile with a background region greater than 95% is eliminated.

3.3.2 Feature extraction/vector embedding:

Once the tiles are extracted from the raw whole slide images and the background regions eliminated, they are now embedded into feature vector representations. This embedding process is done through a pretrained vision transformer called Ctranspath (Wang *et al.*, 2022). Ctranspath is

a foundational model built using the combination of Convolutional Neural Networks (CNN) (Lecun *et al.*, 1998) and a hierarchical Vision Transformer architecture called SwinTransformers as backbone (Liu *et al.*, 2021). This hybrid architecture is then trained on approximately 15 million image patches extracted from 35000 Whole Slide images of multiple cancers (Wang *et al.*, 2022) using Self-Supervised Contrastive Learning (Ericsson *et al.*, 2022). This rigorous training process helps the model learn general-purpose high-quality feature representations from image tiles, capturing spatial and morphological information at a local context (Tile level) while also accounting for information at a global context (slide level). In this scenario, a pretrained model was selected over training a vision transformer from scratch because a foundational model requires a substantially large amount of data (usually in Tens to hundreds of Terabytes) to learn intricate feature representations of the histopathological images, which was not readily available (Dosovitskiy *et al.*, 2021). Hence, by using a pretrained transformer model, it was possible to leverage the robustness of a model that was trained on a large set of data and apply it to the bladder cancer WSIs of this analysis. Therefore, by using the Ctranspath feature extractor, each tile is embedded into a feature vector representation of length 1024 that captures morphological and contextual information of the tile images, which can be used downstream.

3.3.3 Slide level vector extraction:

Although each tile has been embedded into vector feature representations, they still don't describe the image as a whole, as they are extracted individually for each tile (Bilal *et al.*, 2023). In order to capture the Inter-tile relationships, the tile-level feature vectors have to be aggregated together to obtain the slide-level feature representation that captures feature relationships at a global context. This aggregation is done using another pretrained model called CHIEF (Clinical Histopathology Imaging Evaluation Foundation) (Wang *et al.*, 2024). The CHIEF model is a multi-instance learning model that is pretrained using weakly supervised learning. The training process was carried out on approximately 65,000 pan-cancer histopathological images (48 terabytes) from 19 different anatomical sites for aggregating tile-level feature representation to produce Whole Slide level features. This aggregation is done using a gated attention mechanism (Ilse, Tomczak and Welling, 2018). Attention mechanism ensures that the model is trained to focus on interesting regions and suppress irrelevant regions in the whole slide image while highlighting the relevant regions important for a particular task (Schlemper *et al.*, 2019). The CHIEF model, in addition to generating a slide-level feature representation, also produces attention scores for individual tiles. This attention score can then be used to identify the most informative region within a Bladder cancer Image. Here again, a pretrained model was preferred due to the same hurdle of limited data availability to train the model from scratch to learn intricate slide features. Therefore, by opting for a pretrained model that has been weakly trained on thousands of histopathological images that are similar to the Bladder cancer images of this study, helps extract a base set of features that acts as a foundation. This foundation can then be fine-tuned and used for modelling specific to the task (Mai *et al.*, 2024). Figure 3 below shows the top 25 most interesting regions in a Whole Slide Image of Bladder cancer obtained based on CHIEF attention scores.

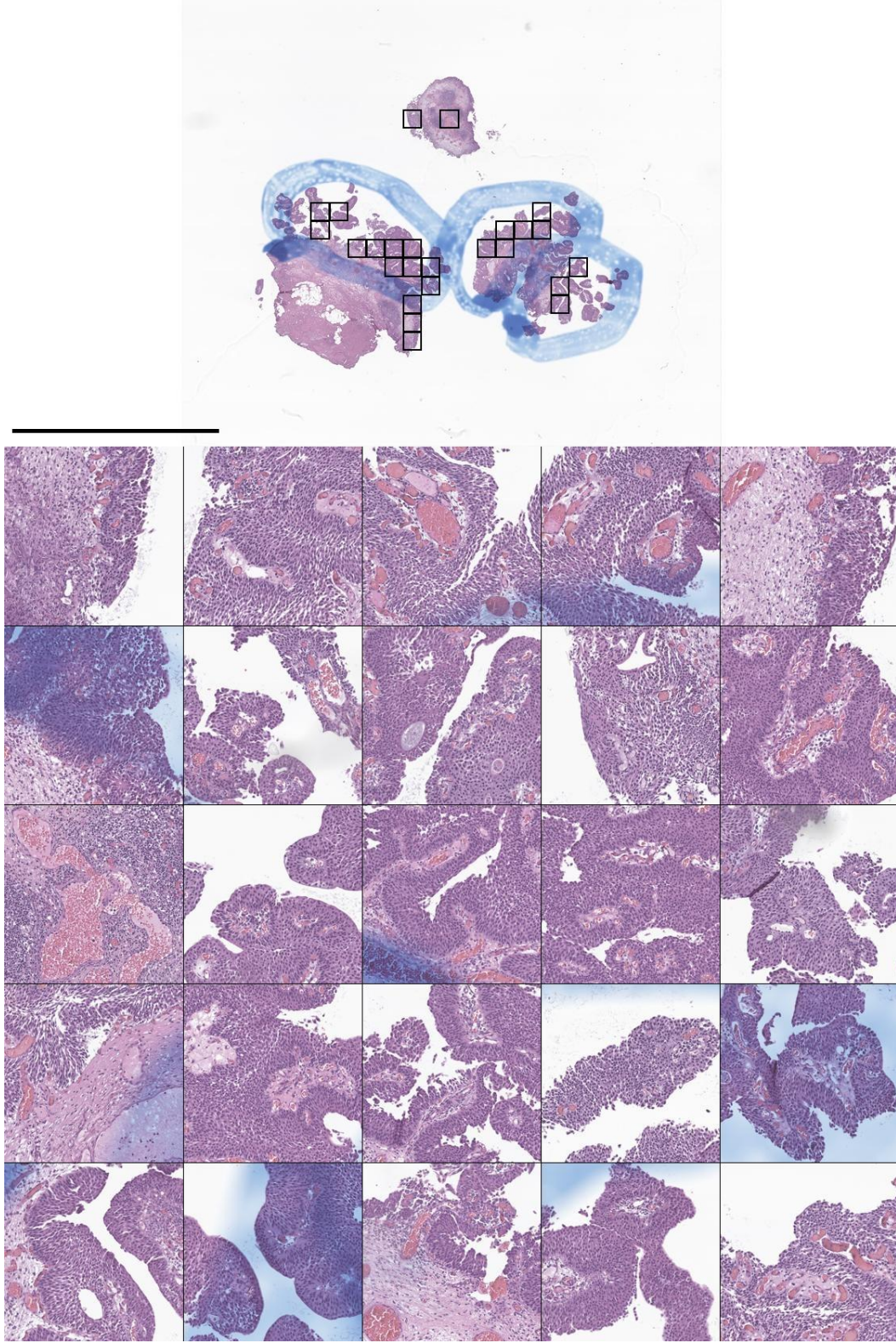


Figure 3: Coordinate location and visualization of most top 25 informative tiles obtained from CHIEF attention scores. The Coordinates of the informative tiles lie mainly within the blue-marked zone, indicating that cancerous regions are considered most informative.

3.4 Task Level:

Now that the slide level feature representations are extracted using the CHIEF aggregator model, these features can be used as baseline features for a number of tasks, such as cancer subtyping, cancer classification, segmentation, image-to-image retrieval, prognosis prediction, etc. (Ochi, Komura and Ishikawa, 2025). In the context of this analysis, the task is predicting the FGFR3 mutation.

3.4.1 Feature visualization:

To understand the features and to examine if any natural clusters separate the features of both the output classes, the feature vectors were projected into a latent space and visualized using dimensionality reduction techniques such as Principal Component Analysis (PCA) (Hotelling, 1933) and Uniform Manifold Approximation and Projection (UMAP) (McInnes *et al.*, 2018).
Figure

3.4.2 Model Building:

Model Architecture:

A Multi Layer Perceptron (MLP) model (Rosenblatt, 1958) that uses back propagation (Rumelhart, Hinton and Williams, 1986) to model the relationship between slide feature vectors and the FGFR3 mutation is developed. The Multi Layer Perceptron consists of an input layer of size 512 (length of the slide feature vector), a configurable set of intermediate layers, and an output layer. Each set in the intermediate layer consists of a fully connected layer, a batch normalization layer (Ioffe and Szegedy, 2015), and a dropout layer (Srivastava *et al.*, 2014) for regularization. The depth of intermediate layers is dynamically determined based on the best fit to the data during training. For example, an MLP of depth one consists of a single set of intermediate layers described above. The activation functions are also chosen dynamically between RELU (Nair and Hinton, 2010) and GELU (Hendrycks and Gimpel, 2023) while training to avoid vanishing gradients. Residual connections are also employed to improve the gradient flow and allow deeper networks (He *et al.*, 2016). The output layer comprises a fully connected layer mapping the final hidden representation to two logits, which are subsequently passed through a softmax function for classification.

Model Training and Evaluation:

The Model was trained on between 75 and 85 percent of the data based on the trial number, and the rest 15 to 25 percent of the data is used as Test data. To find the best hyperparameters (Activation function, intermediate layer depth, etc.) Optuna hyperparameter search algorithm (Akiba *et al.*, 2019), a version of Bayesian optimization (Moćkus, 1975), is used to search the hyperparameter space directionally. 5-fold cross-validation is used within the training data, where the model with selected hyperparameters is trained on 4 folds and evaluated in the last fold. The model with the best hyperparameter set is then trained on the entire training data (75-85%) and

evaluated on the test data (15-25%). The entire experiment is repeated ten times with different train-test splits to ensure consistency in model performance.

3.4.2 Interpretation:

Now that a predictive model is established, it becomes crucial to understand how the model differentiates between the cancer images from the class where the FGFR3 mutation is present and the class with wild types. Therefore, focusing on interpretable modelling becomes essential to understand the predictions (van der Velden *et al.*, 2022). A model-agnostic interpretation strategy is used to understand the image features corresponding to the classes and assess the quality of informative tiles obtained in the foundational level. Here, Interpretative visualization is carried out in two different parts. The first part of the interpretation is focused on identifying the regions in the Bladder cancer image that show high similarity against the top 3 informative tiles of the same image. Cosine similarity was computed for the feature vectors of the tiles obtained using Ctranspath against the top 3 tiles. A Threshold of 0.75 was set for the average similarity for a tile to be considered highly similar. Figures 7 and 8 show the similarity visualization for the top three tiles within a slide image. In the second part of the interpretation, a similar process to part one is carried out, but instead of the similarity check being within an image, it is carried out across the two classes. This is done by selecting 50 random images from a particular class and picking the top 3 informative tiles from each image ($3 \times 50 = 150$ informative tiles). Then a similarity comparison is done against all the other images ($300 - 50 = 250$ images) to see how common these tiles are across both classes. Figures 9 and 10 depict interpretive visualization across classes.

4. Results:

4.1 Top 25 Informative Tiles:

The important tiles extracted from the CHIEF gated attention are visualized in Figure 3 below. Figure 3 shows majority of the informative tiles extracted fall under the pre-marked zone of interest (the blue region). This particular trend was seen in almost all image samples, indicating that the CHIEF model can identify tiles that are more likely to be cancerous. This consistency in informative tiles being located within the marked region of interest demonstrated the validity of the attention mechanism applied and testifies to the model's generalizability.

4.2 Feature Visualization:

When the slide feature vectors were projected into latent space using principal component analysis (PCA), The First and second principal components were able to capture around 50 percent of the total variance in the data. Also, the scatter plot of principal component one and principal component two showed the formation of two natural clusters. Figure 4 depicts that points with no mutation (blue points) tend to be more concentrated around the center, while the points with mutation present (red points) are spread out on all sides.

Similarly, when the features are projected using UMAP with cosine similarity as the distance metric, the difference between the natural clusters is more distinct. Figure 5 shows that the points from the class mutation present are clustered towards the top, while the points from the class with mutation absent are clustered towards the bottom. These consistent natural clusters in both PCA and UMAP show that the features embedded have captured spatial and morphological information from the images that differentiate between classes well.

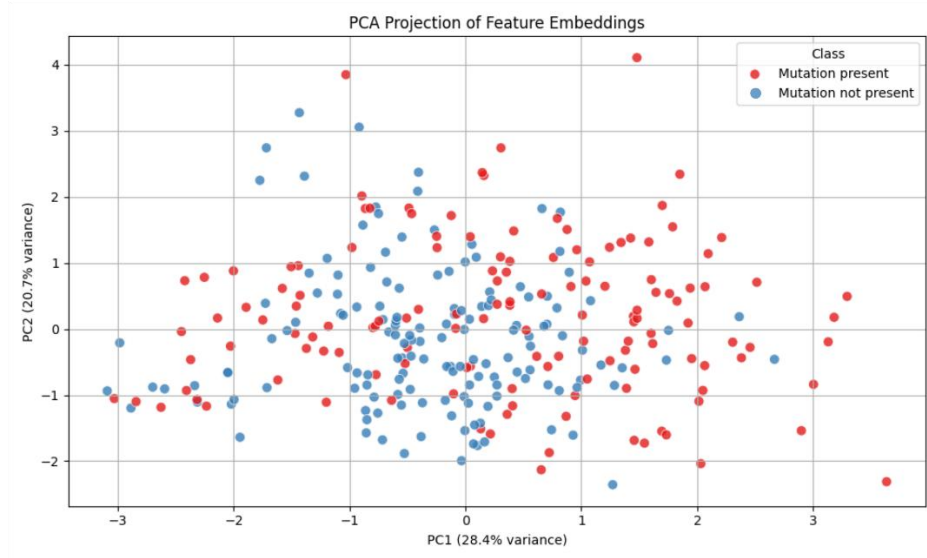


Figure 4: Slide level feature vectors projected into latent space through PCA and PC1 and PC2 plotted against each other.

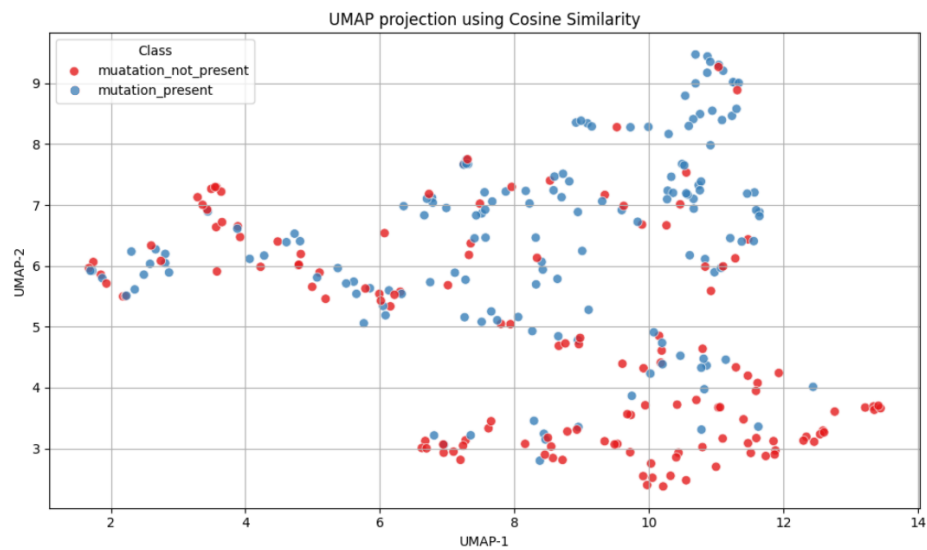


Figure 5: Slide level feature vectors projected into latent space through UMAP with cosine as distance and UMAP1 and UMAP2 plotted against each other.

4.3 Predictive Modelling:

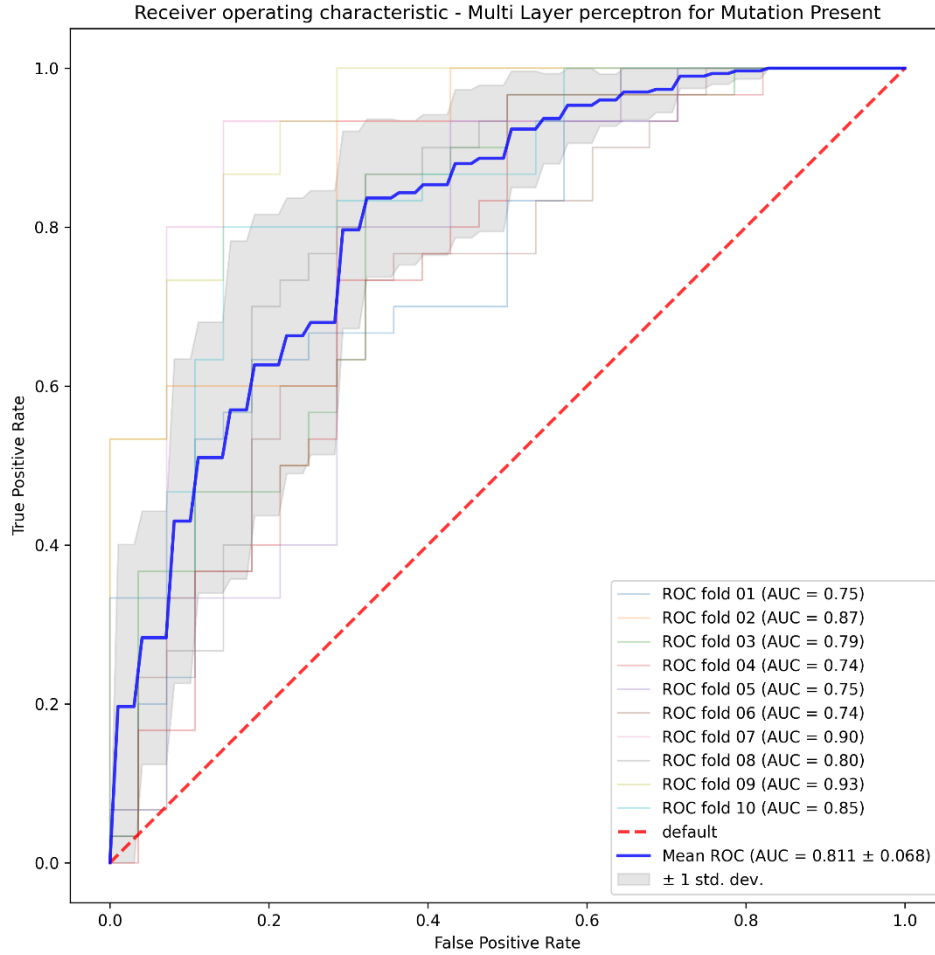


Figure 6: Receiver Operating Characteristics for the Multi-layer perceptron across 10 different trials using changing train and test split.

The Receiver Operating Characteristic (ROC) plot for the MLP model across 10 Trials is shown in Figure 6. The model had a consistent performance with an average AUC of 0.81 and a standard deviation of 0.068. The model performed the best at trial 9 with a peak AUC of 0.93, and the worst at trial 6 with an AUC of 0.74, which still shows consistent performance, indicating that there are visual features in the Whole Slide Images that differentiate the cancer with FGFR3 mutation present and FGFR3 mutation not present.

4.4 Interpretive visualization:

For the first part of the visualization, where the cosine similarity against the top three tiles of the Whole slide image was estimated, it was observed that the regions within the pre-marked region of interest were highly similar (average similarity greater than 0.75) to the top 3 informative tiles obtained from CHIEF attention. Figures 7 and 8 show the similarity visualization against the top 3 informative tiles within a slide, where the marked regions exhibit high similarity with the

informative tiles obtained from CHIEF attention. The top 3 informative tiles are marked with square blocks.

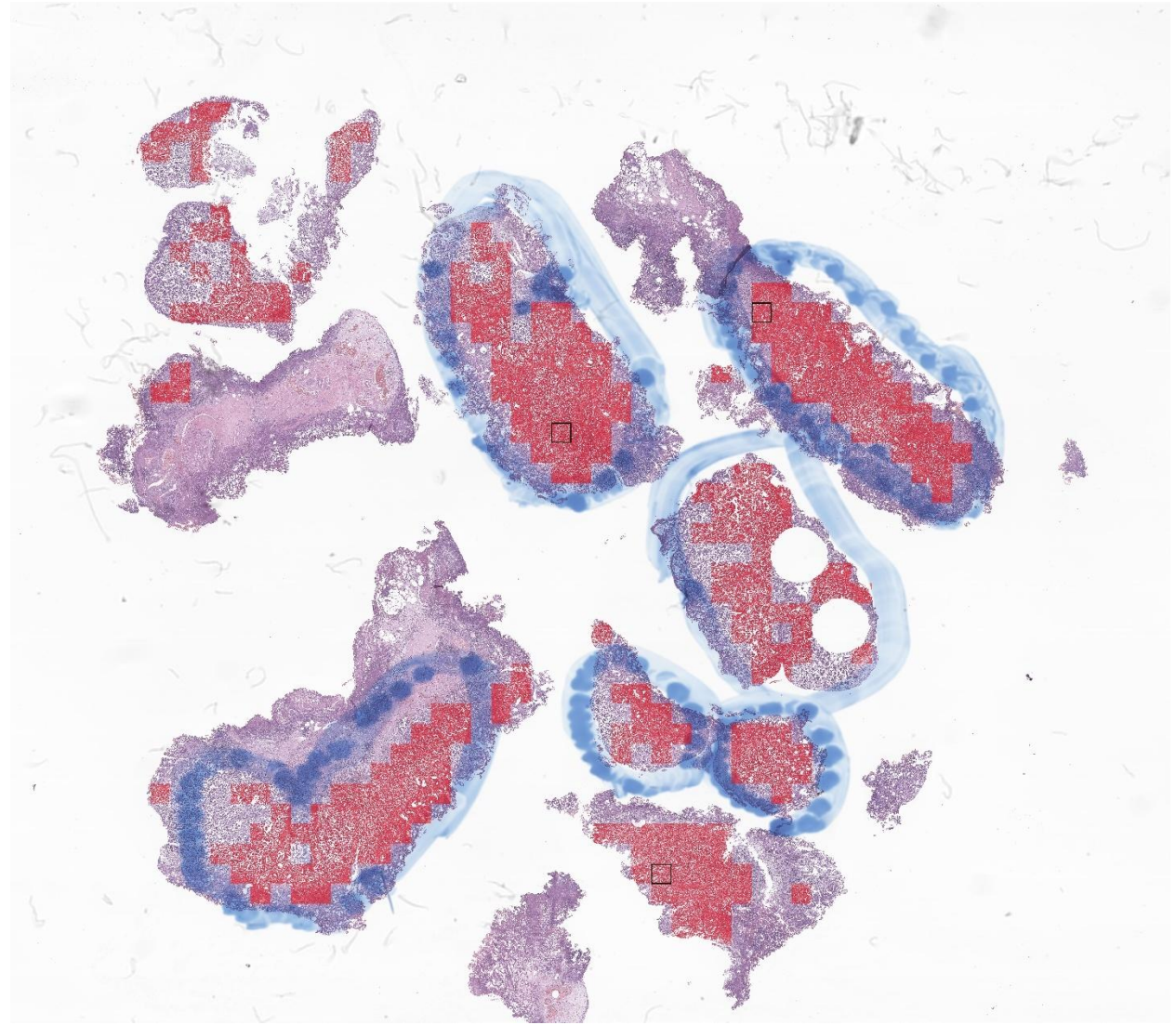


Figure 7: Example 1 of Similarity visualization of the top three informative tiles (Marked in square blocks) against the rest of the tiles using cosine similarity to identify similar tiles within the Bladder cancer image and understand their frequency and location. Tiles marked in red have average cosine similarity greater than 0.75.

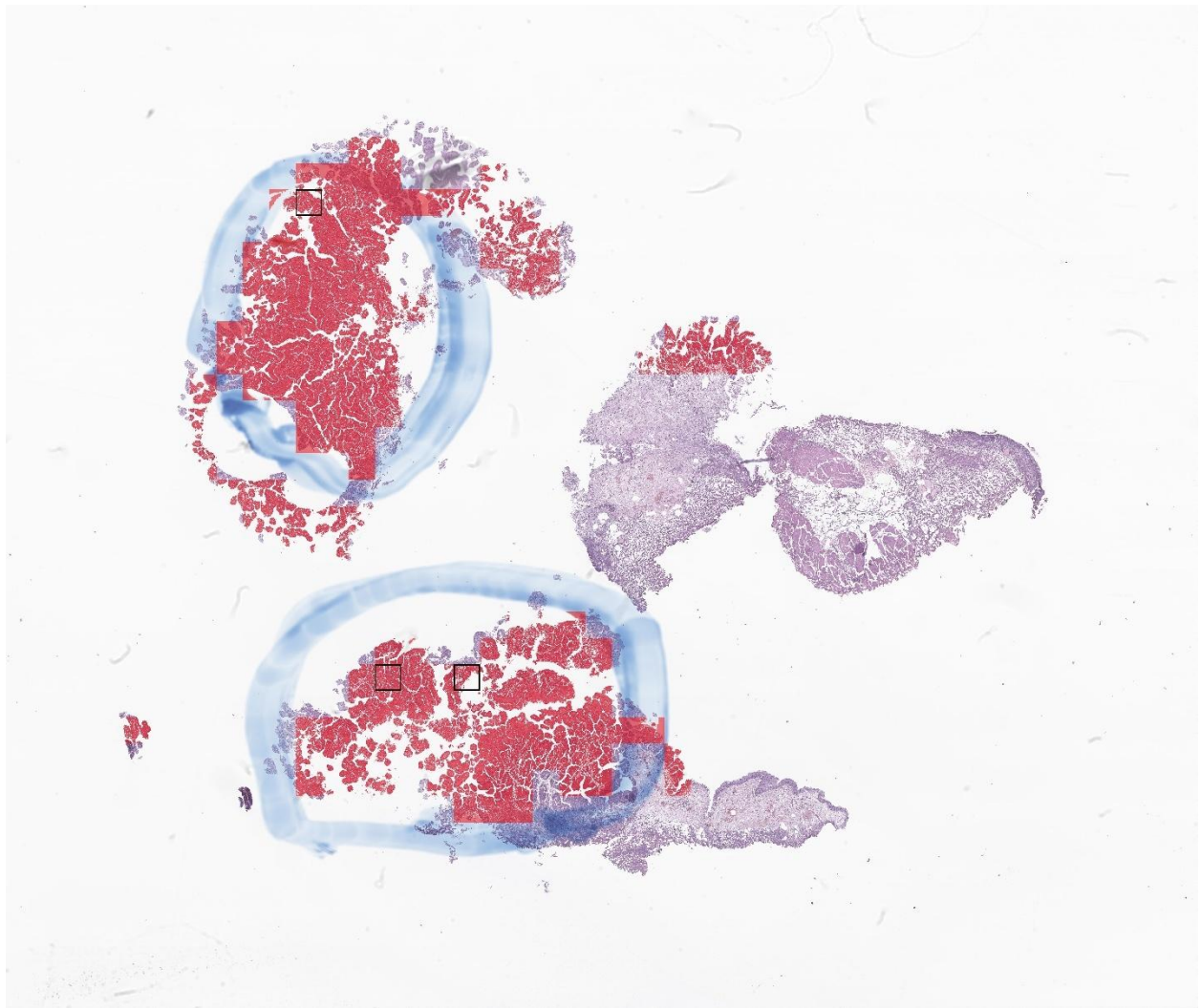
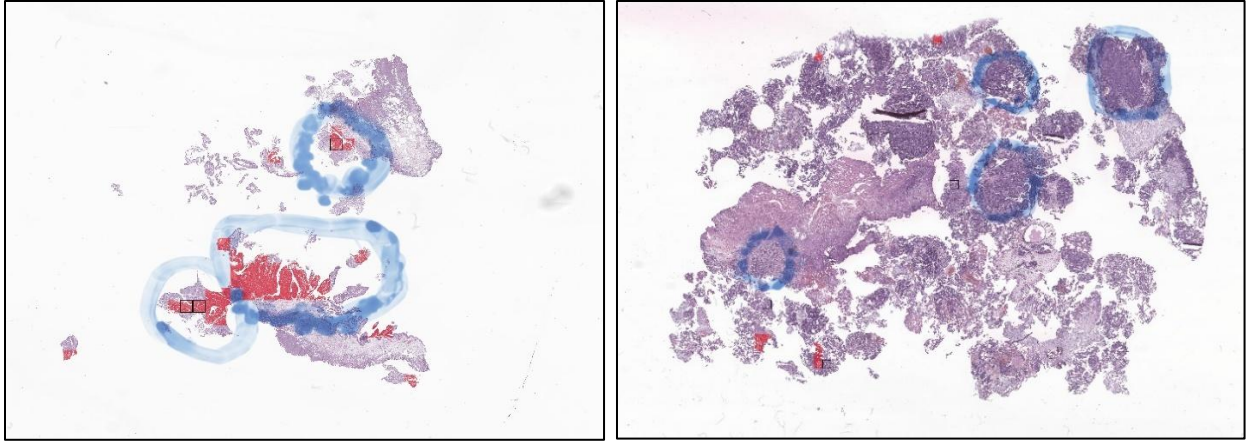


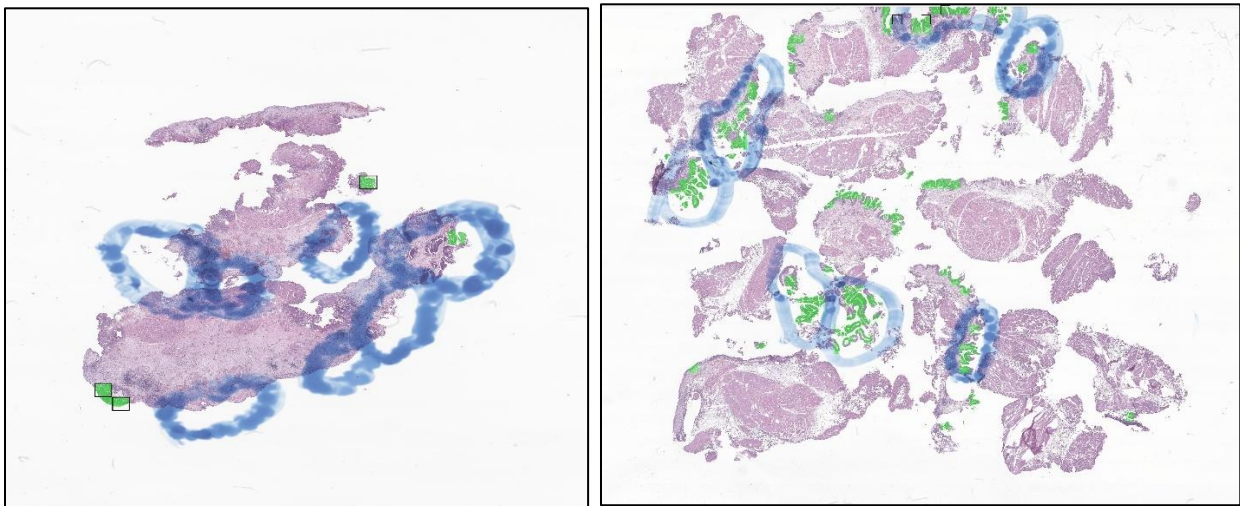
Figure 8: Example 2 of Similarity visualization of the top three informative tiles (Marked in square blocks) against the rest of the tiles using cosine similarity to identify similar tiles within the Bladder cancer image and understand their frequency and location. Tiles marked in red have average cosine similarity greater than 0.75.

In the second part of the visualization, the top 3 tiles from all 50 reference images of one reference class (can be either mutation present or wild type) are compared against both classes (the remaining 250 images) as shown in Figures 9 and 10 below. The general trend observed here is that the tiles from the reference class seem to be more similar to the images from the same class. For example, suppose the reference tiles are from a mutation present class. In that case, they seem to be more similar to the rest of the images from the mutation present class, while less common in the wild type, thereby showing some difference in cancer between the two classes.



Reference class: FGFR3 Mutation present. **Left:** FGFR3 Mutation present. **Right:** Wild Type

Figure 9: Example 1 of similarity visualization of top tiles from 50 reference images of one class against both classes to examine intra-class heterogeneity, images of the mutation present class, and wild types.



Reference class: Wild Type. **Left:** FGFR3 Mutation present. **Right:** Wild Type

Figure 10: Example 2 of similarity visualization of top tiles from 50 reference images of one class against both classes to examine intra-class heterogeneity, images of the mutation present class, and wild types.

5. Discussion:

Different targeted therapies for bladder cancer are being developed to focus on the FGFR3 signaling pathway. One notable example is the recently approved drug Erdafitinib, which inhibits tyrosine kinase targeting FGFR3 signaling pathways (Loriot *et al.*, 2019; Noeraparast, Krajina, Pichler, Niedersüß-Beke, Shahrokh F. Shariat, *et al.*, 2024). Targeted therapies for different molecular alterations have become the focus of treatment in bladder cancer. Still, they are held

back due to the genomic screening practices that take a lot of time. For example, both FDA and EMA approval of Erdafitinib is bound to proving the existence of the FGFR3 mutation in the first place.

In this analysis, we have created a deep learning-based classification model that predicts the FGFR3 mutation status in bladder cancer using histopathological images. The model's performance, with an Area under the curve averaging at 0.81, shows that there is some visual representation in the histopathological images that helps the model differentiate between the cancer with FGFR3 mutation present and the mutation not present.

In the feature embedding stage, when feature vectors were extracted from the whole slide images, the attention mechanism mainly prioritized the cancerous regions. This was backed by locating the highly informative tiles selected by the model during the aggregation phase, as shown in Figure 3. The results from the interpretive visualization part, represented in Figures 7 and 8, also indicated the same result, where the informative tiles were very similar to the tiles within the pre-marked region of interest. This suggests that the cancer region may play a significant role in determining the mutation status in bladder cancer whole-slide images.

In interpretive visualization, where the intra-class differences were examined, the major regions that differ across both classes were the cancerous regions, as shown in Figure 9 and 10. This backs the original hypothesis of intra-cancerous differences between the two classes, where Non-Muscle invasive bladder (NMIBC) is often triggered by a FGFR3 mutation and very rarely seen in Muscle invasive bladder cancer (Noeraparast, Krajina, Pichler, Niedersüß-Beke, Shahrokh F Shariat, *et al.*, 2024). (Loeffler *et al.*, 2022a) reported similar observations of intra-tumoral heterogeneity based on the mutation status, where the papillary morphology played the most pivotal role in differentiating the bladder cancers with and without the mutation. In addition to the papillary morphology, the classification also seemed to be based on “polar epithelial cell sheets, roundish areas, and densely packed, but homogenous, regular nuclei.” (Loeffler *et al.*, 2022b, p. 5).

A major limitation of our approach lies in the tradeoff between representation power and interpretability. While transforming WSIs into feature vectors enables deep learning at scale, it obscures the local, feature-level contributions driving classification decisions. Moreover, our dataset lacked labels for tumor subtypes, preventing a more nuanced analysis of subtype-mutation relationships. Since FGFR3 mutations are more commonly associated with non-muscle-invasive bladder cancer (NMIBC), incorporating subtype annotations could help uncover whether the model captures indirect associations mediated by histological subtype.

Future work should focus on improving model interpretability, e.g., via concept-based explanations or class activation maps, and on validating the biological relevance of learned features. Incorporating labeled subtype information will also enable a more comprehensive assessment of intra-tumor heterogeneity and the causal or correlative link between FGFR3 mutation and cancer morphology.

6. Conclusion:

From this study, it can be concluded that there are visual features present in histopathological images of bladder cancer that can help in determining the FGFR3 mutation status. Timely identification of the FGFR3 mutation helps in identifying the bladder cancer subtypes quickly to focus on targeted treatment strategies, which promote better prognosis and improve patient outcomes. In this study, we have trained a deep learning model that can identify FGFR3 mutation in bladder cancer using Whole Slide images. The ability to predict mutation status with high confidence demonstrates that deep learning and histopathological images can serve as a pre-screening tool in identifying the FGFR3 mutation, thereby reducing turnaround times. This can further help the healthcare system in saving costs and effort spent on sequencing studies to identify the FGFR3 mutation status, while also improving the quality of outcomes.

Code availability and supplementary:

The dataset is not available publicly. The code and supplementary figures obtained from the analysis can be found at [Github repository](#).

7. References:

- Akiba, T. *et al.* (2019) ‘Optuna: A Next-generation Hyperparameter Optimization Framework’, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: Association for Computing Machinery (KDD ’19), pp. 2623–2631. Available at: <https://doi.org/10.1145/3292500.3330701>.
- Alexandrov, L.B. *et al.* (2013) ‘Signatures of mutational processes in human cancer’, *Nature*, 500(7463), pp. 415–421. Available at: <https://doi.org/10.1038/nature12477>.
- Bannier, P.-A. *et al.* (2024) ‘AI allows pre-screening of FGFR3 mutational status using routine histology slides of muscle-invasive bladder cancer’, *Nature Communications*, 15(1), p. 10914. Available at: <https://doi.org/10.1038/s41467-024-55331-6>.
- Bilal, M. *et al.* (2023) ‘An aggregation of aggregation methods in computational pathology’, *Medical Image Analysis*, 88, p. 102885. Available at: <https://doi.org/10.1016/j.media.2023.102885>.
- Chen, M. *et al.* (2020) ‘Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning’, *NPJ precision oncology*, 4, p. 14. Available at: <https://doi.org/10.1038/s41698-020-0120-3>.
- Dosovitskiy, A. *et al.* (2021) ‘AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE’.
- Dyrskjot, L. *et al.* (2023) ‘Bladder cancer’, *Nature Reviews Disease Primers*, 9(1), p. 58. Available at: <https://doi.org/10.1038/s41572-023-00468-9>.

El Nahhas, O.S.M. *et al.* (2025) ‘From whole-slide image to biomarker prediction: end-to-end weakly supervised deep learning in computational pathology’, *Nature Protocols*, 20(1), pp. 293–316. Available at: <https://doi.org/10.1038/s41596-024-01047-2>.

Ericsson, L. *et al.* (2022) ‘Self-Supervised Representation Learning: Introduction, advances, and challenges’, *IEEE Signal Processing Magazine*, 39(3), pp. 42–62. Available at: <https://doi.org/10.1109/MSP.2021.3134634>.

Faghani, S. *et al.* (2023) ‘Optimizing Storage and Computational Efficiency: An Efficient Algorithm for Whole Slide Image Size Reduction’, *Mayo Clinic Proceedings: Digital Health*, 1(3), pp. 419–424. Available at: <https://doi.org/10.1016/j.mcpdig.2023.06.010>.

He, K. *et al.* (2016) ‘Deep Residual Learning for Image Recognition’, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. Available at: <https://doi.org/10.1109/CVPR.2016.90>.

Hendrycks, D. and Gimpel, K. (2023) ‘Gaussian Error Linear Units (GELUs)’. arXiv. Available at: <https://doi.org/10.48550/arXiv.1606.08415>.

Hotelling, H. (1933) ‘Analysis of a complex of statistical variables into principal components’, *Journal of Educational Psychology*, 24(6), pp. 417–441. Available at: <https://doi.org/10.1037/h0071325>.

Ilse, M., Tomczak, J. and Welling, M. (2018) ‘Attention-based Deep Multiple Instance Learning’, in *Proceedings of the 35th International Conference on Machine Learning. International Conference on Machine Learning*, PMLR, pp. 2127–2136. Available at: <https://proceedings.mlr.press/v80/ilse18a.html> (Accessed: 24 July 2025).

Ioffe, S. and Szegedy, C. (2015) ‘Batch normalization: accelerating deep network training by reducing internal covariate shift’, in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. Lille, France: JMLR.org (ICML’15), pp. 448–456.

Lawrence, M.S. *et al.* (2013) ‘Mutational heterogeneity in cancer and the search for new cancer-associated genes’, *Nature*, 499(7457), pp. 214–218. Available at: <https://doi.org/10.1038/nature12213>.

Lenis, A.T. *et al.* (2020) ‘Bladder Cancer: A Review’, *JAMA*, 324(19), pp. 1980–1991. Available at: <https://doi.org/10.1001/jama.2020.17598>.

Liu, Z. *et al.* (2021) ‘Swin Transformer: Hierarchical Vision Transformer using Shifted Windows’, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9992–10002. Available at: <https://doi.org/10.1109/ICCV48922.2021.00986>.

Loeffler, C.M.L. *et al.* (2022a) ‘Artificial Intelligence–based Detection of FGFR3 Mutational Status Directly from Routine Histology in Bladder Cancer: A Possible Preselection for Molecular

Testing?', *European Urology Focus*, 8(2), pp. 472–479. Available at: <https://doi.org/10.1016/j.euf.2021.04.007>.

Loeffler, C.M.L. *et al.* (2022b) 'Artificial Intelligence–based Detection of FGFR3 Mutational Status Directly from Routine Histology in Bladder Cancer: A Possible Preselection for Molecular Testing?', *European Urology Focus*, 8(2), pp. 472–479. Available at: <https://doi.org/10.1016/j.euf.2021.04.007>.

Loriot, Y. *et al.* (2019) 'Erdafitinib in Locally Advanced or Metastatic Urothelial Carcinoma', *The New England Journal of Medicine*, 381(4), pp. 338–348. Available at: <https://doi.org/10.1056/NEJMoa1817323>.

Lu, M.Y. *et al.* (2024) 'A visual-language foundation model for computational pathology', *Nature Medicine*, 30(3), pp. 863–874. Available at: <https://doi.org/10.1038/s41591-024-02856-4>.

Mai, G. *et al.* (2024) 'On the Opportunities and Challenges of Foundation Models for GeoAI (Vision Paper)', *ACM Trans. Spatial Algorithms Syst.*, 10(2), p. 11:1-11:46. Available at: <https://doi.org/10.1145/3653070>.

McInnes, L. *et al.* (2018) 'UMAP: Uniform Manifold Approximation and Projection', *Journal of Open Source Software*, 3(29), p. 861. Available at: <https://doi.org/10.21105/joss.00861>.

Močkus, J. (1975) 'On bayesian methods for seeking the extremum', in G.I. Marchuk (ed.) *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974*. Berlin, Heidelberg: Springer, pp. 400–404. Available at: https://doi.org/10.1007/3-540-07165-2_55.

Nair, V. and Hinton, G.E. (no date) 'Rectified Linear Units Improve Restricted Boltzmann Machines'.

Noeraparast, M., Krajina, K., Pichler, R., Niedersüß-Beke, D., Shariat, Shahrokh F., *et al.* (2024) 'FGFR3 alterations in bladder cancer: Sensitivity and resistance to targeted therapies', *Cancer Communications (London, England)*, 44(10), pp. 1189–1208. Available at: <https://doi.org/10.1002/cac2.12602>.

Noeraparast, M., Krajina, K., Pichler, R., Niedersüß-Beke, D., Shariat, Shahrokh F, *et al.* (2024) 'FGFR3 alterations in bladder cancer: Sensitivity and resistance to targeted therapies', *Cancer Communications*, 44(10), pp. 1189–1208. Available at: <https://doi.org/10.1002/cac2.12602>.

Ochi, M., Komura, D. and Ishikawa, S. (2025) 'Pathology Foundation Models', *JMA Journal*, 8(1), pp. 121–130. Available at: <https://doi.org/10.31662/jmaj.2024-0206>.

Otsu, N. (1979) 'A Threshold Selection Method from Gray-Level Histograms', *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), pp. 62–66. Available at: <https://doi.org/10.1109/TSMC.1979.4310076>.

Rosenblatt, F. (1958) 'The perceptron: A probabilistic model for information storage and organization in the brain', *Psychological Review*, 65(6), pp. 386–408. Available at: <https://doi.org/10.1037/h0042519>.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) ‘Learning representations by back-propagating errors’, *Nature*, 323(6088), pp. 533–536. Available at: <https://doi.org/10.1038/323533a0>.

Schlemper, J. *et al.* (2019) ‘Attention gated networks: Learning to leverage salient regions in medical images’, *Medical Image Analysis*, 53, pp. 197–207. Available at: <https://doi.org/10.1016/j.media.2019.01.012>.

Sirinukunwattana, K. *et al.* (2021) ‘Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning’, *Gut*, 70(3), pp. 544–554. Available at: <https://doi.org/10.1136/gutjnl-2019-319866>.

Srivastava, N. *et al.* (2014) ‘Dropout: A Simple Way to Prevent Neural Networks from Overfitting’, *Journal of Machine Learning Research*, 15(56), pp. 1929–1958.

Szegedy, C. *et al.* (2016) ‘Rethinking the Inception Architecture for Computer Vision’, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826. Available at: <https://doi.org/10.1109/CVPR.2016.308>.

Tran, L. *et al.* (2021) ‘Advances in bladder cancer biology and therapy’, *Nature Reviews Cancer*, 21(2), pp. 104–121. Available at: <https://doi.org/10.1038/s41568-020-00313-1>.

Vaswani, A. *et al.* (2017) ‘Attention is all you need’, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc. (NIPS’17), pp. 6000–6010.

van der Velden, B.H.M. *et al.* (2022) ‘Explainable artificial intelligence (XAI) in deep learning-based medical image analysis’, *Medical Image Analysis*, 79, p. 102470. Available at: <https://doi.org/10.1016/j.media.2022.102470>.

Wang, R. *et al.* (2025) ‘Deep learning for predicting prognostic consensus molecular subtypes in cervical cancer from histology images’, *npj Precision Oncology*, 9(1), p. 11. Available at: <https://doi.org/10.1038/s41698-024-00778-5>.

Wang, X. *et al.* (2022) ‘Transformer-based unsupervised contrastive learning for histopathological image classification’, *Medical Image Analysis*, 81, p. 102559. Available at: <https://doi.org/10.1016/j.media.2022.102559>.

Wang, X. *et al.* (2024) ‘A pathology foundation model for cancer diagnosis and prognosis prediction’, *Nature*, 634(8035), pp. 970–978. Available at: <https://doi.org/10.1038/s41586-024-07894-z>.

Zarella, M.D. *et al.* (2018) ‘A Practical Guide to Whole Slide Imaging: A White Paper From the Digital Pathology Association’, *Archives of Pathology & Laboratory Medicine*, 143(2), pp. 222–234. Available at: <https://doi.org/10.5858/arpa.2018-0343-RA>.

