biliary cancer, extrahepatic bile duct cancer, intrahepatic bile duct cancer, bladder cancer, urinary bladder cancer, osteosarcoma, brain tumor, brain stem glioma, breast cancer (including triple negative breast cancer), cervical cancer, colon cancer, colorectal cancer, lymphoma, endometrial cancer, esophageal cancer, gastric (stomach) cancer, head and neck cancer, hepatocellular (liver) cancer, kidney cancer, renal cancer, lung cancer, melanoma, cancer of the tongue, oral cancer, ovarian cancer, pancreatic cancer, prostate cancer, uterine cancer, testicular cancer, and vaginal cancer.

[0016] In some examples, the imaging-based biomarker prediction systems are formed of deep learning frameworks having a multiscale configuration designed to perform classification on (labeled or unlabeled) histopathology images using classifiers trained to classify tiles of received histopathology images. In some examples, the multiscale configurations contain tile-level tissue classifiers, i.e., classifiers trained using tile-based deep learning training. In some examples, the multiscale configurations contain pixel-level cell classifiers and cell segmentation models. In some examples, the classifications form the tile-level tissue classifiers and from the pixel-level cell classifiers are analyzed to predict biomarker status in the histopathology image. In yet some examples, the multiscale configurations contain tile-level biomarker classifiers.

[0017] In some examples, the imaging-based biomarker prediction systems are formed of deep learning frameworks having a single-scale configuration designed to perform classifications on (labeled or unlabeled) histopathology images using classifiers trained using multiple instance learning (MIL) techniques. In some examples, the single-scale configurations contain slide-level classifiers trained using gene sequencing data, such as RNA sequencing data. That is, slide-level classifiers are trained using RNA sequence data to develop image-based classifiers capable of predicting biomarker status in histopathology images.

[0018] According to a first aspect of the present invention, there is provided a computer-implemented method of identifying biomarkers in digital image of a hematoxylin and eosin (H&E) stained slide of target tissue, the method comprises: receiving the digital image to an image-based biomarker prediction system having one or more processors; performing an image tiling process, using the one or more processors, on the digital image by separating the digital image into a plurality of tile images, where each of the plurality of tile images contains a different portion of the digital image; applying, using the one or more processors, the plurality of tile images to a multiscale deep learning framework comprising: (i) one or more trained deep learning multiscale classifier models, each trained deep learning multiscale classifier models being trained to classify a different tissue classification for each tile image and to determine a tissue classification for each of the plurality of tile images, using the multiscale deep learning framework, and (ii) a trained cell segmentation model that processes pixels of the digital image; identifying, using the one or more processors, cells within

the digital image using the trained cell segmentation model; and from the tissue classification determined for each tile image and from the identified cells within the digital image, identifying a predicted presence of one or more biomarkers associated with the digital image.

[0019] In accordance with another example, a computer-implemented method of identifying biomarkers in a digital image of a hematoxylin and eosin (H&E) stained slide of target tissue, the method comprises: receiving a molecular training dataset for a plurality of training tissue samples, the molecular training dataset comprising RNA transcriptome counts from sequencing of a substantially similar sample associated with each training tissue sample; performing a clustering process on the molecular training dataset to identify one or more molecular data subsets each corresponding to a different biomarker; for each of the one or more molecular data subsets receiving a plurality of digital images of H&E stained training slides of training tissue samples corresponding to the respective biomarker to an image-based biomarker prediction system having one or more processors; generating, using the one or more processors, for each of the one or more molecular data subsets a trained image-based biomarker classifier model, based on the plurality of digital images of the H&E stained training slides; receiving, using the one or more processors, a subsequent digital image of an H&E stained slide of a subsequent tissue sample; and apply, using the one or more processors, the subsequent digital image to the trained image-based biomarker classifier model and identifying a predicted presence of one or more biomarkers for the subsequent tissue sample.

[0020] In accordance with another example, a computer-implemented method of identifying biomarkers in a digital image of a hematoxylin and eosin (H&E) stained slide of target tissue, the method comprises: receiving the digital image to an image-based biomarker prediction system having one or more processors; separating, using the one or more processors, the digital image into a plurality of tile images, where each of the plurality of tile images contains a different portion of the digital image; applying, using the one or more processors, the plurality of tile images to a deep learning framework comprising one or more trained biomarker classification models, each trained biomarker classification model being trained to classify a different biomarker; predicting, using the one or more processors, a biomarker classification for each of the plurality of tile images using the one or more trained biomarker classification models; from the predicted biomarker classifications of each of the tile images, determining a predicted presence of one or more biomarkers in the target tissue; and generating a report containing the digital image and a digital overlay visualizing the predicted presence of the one or more biomarkers.

[0021] In some examples, the deep learning framework comprises a multiscale deep learning framework.

[0047] In some examples, the pathology slide scanner system contains an image-based, adversarial trained, and/or a microsatellite instability (MSI) prediction model.

[0048] In some examples, generating the report containing the digital image and the digital overlay comprises generating the digital overlay to include an overlay element identifying tumor content of the digital image or tumor percentage of the digital image.

[0049] According to a second aspect of the present invention, there is provided a computing device configured to identifying biomarkers in digital image of a Hematoxylin and eosin (H&E) stained slide of target tissue, the computing device comprises: one or more memories; and one or more processors configured to, receive the digital image; perform an image tiling process on the digital image by separating the digital image into a plurality of tile images, where each of the plurality of tile images contains a different portion of the digital image; apply the plurality of tile images to a multiscale deep learning framework comprising: (i) one or more trained deep learning multiscale classifier models, each trained deep learning multiscale classifier models being trained to classify a different tissue classification for each tile image and to determine a tissue classification for each of the plurality of tile images, using the multiscale deep learning framework, and (ii) a trained cell segmentation model that processes pixels of the digital image; identify cells within the digital image using the trained cell segmentation model; and from the tissue classification determined for each tile image and from the identified cells within the digital image, identify a predicted presence of one or more biomarkers associated with the digital image.

[0050] According to a third aspect of the present invention, there is provided a computing device configured to identifying biomarkers in digital image of a Hematoxylin and eosin (H&E) stained slide of target tissue, the computing device comprises: one or more memories; and one or more processors configured to, receive a molecular training dataset for a plurality of training tissue samples, the molecular training dataset comprising RNA transcriptome counts from sequencing of a substantially similar sample associated with each training tissue sample; perform a clustering process on the molecular training dataset to identify one or more molecular data subsets each corresponding to a different biomarker; for each of the one or more molecular data subsets, receive a plurality of digital images of H&E stained training slides of training tissue samples corresponding to the respective biomarker to an image-based biomarker prediction system having one or more processors; generate, for each of the one or more molecular data subsets, a trained image-based biomarker classifier model, based on the plurality of digital images of the H&E stained training slides and a trained cell segmentation model that processes pixels of the digital image; receive a subsequent digital image of an H&E stained slide of a subsequent tissue sample; and apply the subsequent digital image to the trained image-based biomarker classifier model and the trained cell segmentation model to identify a predicted presence of one or more biomarkers for the subsequent tissue sample.

What is Claimed:

1. A computer-implemented method of identifying biomarkers in digital image of a hematoxylin and eosin (H&E) stained slide of target tissue, the method comprising:

receiving the digital image to an image-based biomarker prediction system having one or more processors;

performing an image tiling process, using the one or more processors, on the digital image by separating the digital image into a plurality of tile images, where each of the plurality of tile images contains a different portion of the digital image;

applying, using the one or more processors, the plurality of tile images to a multiscale deep learning framework comprising:

(i) one or more trained deep learning multiscale classifier models, each trained deep learning multiscale classifier models being trained to classify a different tissue classification for each tile image and to determine a tissue classification for each of the plurality of tile images, using the multiscale deep learning framework, and

(ii) a trained cell segmentation model that processes pixels of the digital image;

identifying, using the one or more processors, cells within the digital image using the trained cell segmentation model; and

from the tissue classification determined for each tile image and from the identified cells within the digital image, identifying a predicted presence of one or more biomarkers associated with the digital image.

2. The method of claim 1, wherein performing the image tiling process on the digital image comprises applying a tiling mask to the digital image, to separate the digital image into the plurality of tile images.

3. The method of claim 2, wherein the tiling mask comprises rectangular tiles of the same size.

4. The method of claim 2, wherein the tiling mask comprises rectangular tiles of differing size.

5. The method of claim 2, wherein the tiling mask comprises tiles characterized by topology and/or morphology of pixels or groups of pixels.

6. The method of claim 1, wherein receiving the digital image comprises, using the one or more processors:

retrieving the digital image at a first image resolution;

downsampling the digital image to a second image resolution;

performing a brightness normalization on pixels in the digital image; and

removing non-tissue objects from the digital image.

7. The method of claim 1, further comprising training the one or more trained deep learning multiscale classifier models by:

receiving, at the multiscale deep learning framework, a plurality of H&E slide training images from a training images dataset, each H&E slide training image having a label corresponding to a biomarker to be trained;

performing tile-based tissue classification analysis on each of the H&E slide training images;

performing a pixel-based cell segmentation analysis on each of the H&E slide training images;

optionally performing a tile-based biomarker classification analysis on each of the H&E slide training images; and

in response, generating the one or more trained deep learning multiscale classifier models.

8. The method of claim 7, wherein each H&E slide training image comprises a plurality of tile images each having a tile-level label.

9. The method of claim 7, further comprising, for each H&E slide training image, imputing a tile-level label for each of a plurality of tile images of the H&E slide training image.

10. The method of claim 7, further comprising:

for each H&E slide training image, performing a tile selection process that infers a class status for each tile image in the H&E slide training image;

based on inferred class status, discarding tile images not corresponding to a desired class, before performing the tile-based tissue classification analysis on each of the H&E slide training images, such that the tile-based tissue classification analysis is performed on only selected tile images of the H&E slide training image.

11. The method of claim 1, further comprising training the one or more trained deep learning multiscale classifier models by:

receiving a molecular training dataset for a plurality of training tissue samples, the molecular training dataset comprising RNA transcriptome counts from sequencing of a substantially similar sample associated with each training tissue sample;

performing a clustering process on the molecular training dataset to identify one or more molecular data subsets each corresponding to a different biomarker;

for each of the one or more molecular data subsets identifying a plurality of digital images of H&E stained training slides of training tissue samples corresponding to the respective biomarker to an image-based biomarker prediction system having one or more processors;

performing tile-based tissue classification analysis on each of the H&E slide training images;

performing a pixel-based cell segmentation analysis on each of the H&E slide training images;

optionally performing a tile-based biomarker classification analysis on each of the H&E slide training images; and

in response, generating the one or more trained deep learning multiscale classifier models.

12. The method of claim 1, wherein the one of the one or more trained deep learning multiscale classifier models are each configured as a tile-resolution Fully Convolutional Network (FCN) classification model.

13. The method of claim 1, wherein the trained deep learning multiscale classifier models are each trained to classify tissue as corresponding to a classification selected from the group consisting of tumor, stroma, normal, lymphocyte, fat, muscle, blood vessel, immune cluster, necrosis, hyperplasia/dysplasia, and red blood cells.

14. The method of claim 1, wherein identifying cells within the digital image tile using the trained cell segmentation model comprises:

applying, using the one or more processors, each of the plurality of tile images to the cell segmentation model and, for each tile, assigning a cell classification to one or more pixels within the tile image.

15. The method of claim 14, wherein assigning the cell classification to one or more pixels within the tile image comprises:

identifying, using the one or more processors, the one or more pixels as a cell interior, a cell border, or a cell exterior and classifying the one or more pixels as the cell interior, the cell border, or the cell exterior.

16. The method of claim 1, wherein the trained cell segmentation model is a pixel-resolution three-dimensional UNet classification model trained to classify a cell interior, a cell border, and a cell exterior.

17. The method of claim 1, wherein identifying cells within the digital image tile using the trained cell segmentation model comprises:

applying, using the one or more processors, each of the plurality of tile images to the cell segmentation model; and

performing, using the one or more processors, a registration on segmented cells in each of the tile images by determining a cell border of each cell, determining a centroid of each cell, and shifting coordinates of centroids to a universal coordinate space for the digital image.

18. The method of claim 1, wherein the trained cell segmentation model is trained using a set of H&E slide training images annotated identifying cell borders, cell interiors, and cell exteriors.

19. The method of claim 1, wherein the digital image is an unlabeled digital image.

20. The method of claim 1, wherein the digital image is a tile-level labeled image or a slide-level labeled image.

21. The method of claim 1, wherein the one or more biomarkers are selected from the group consisting of tumor-infiltrating lymphocytes (TILs), nucleus-to-cytyoplasm (NC) ratio, ploidy, signet ring morphology, and programmed death-ligand 1 (PD-L1).

22. The method of claim 1, wherein the one or more biomarkers is TILs, the method further comprising:

identifying, using the one or more processors, lymphocyte cells within the digital image using a trained lymphocyte segmentation model, by integrating, using the one or more processors, the cell borders identified using the cell segmentation model with the lymphocytes identified using the lymphocyte segmentation model and generating, using the one or more processors, a nested classification of each cell.

23. The method of claim 1, wherein the one or more biomarkers is TILs, the method further comprising:

identifying, using the one or more processors, lymphocyte cells within the digital image using a trained lymphocyte segmentation model, by applying each of a plurality of tile images to a trained lymphocyte segmentation model and, for each tile image, assigning a lymphocyte classification to one or more pixels within the tile image.

24. The method of claim 1, wherein the one or more biomarkers is TILs, the method further comprising identifying, using the one or more processors, lymphocyte cells within the digital image using a trained lymphocyte segmentation model, wherein the lymphocyte segmentation model is a pixel-resolution two-dimensional UNet classification model trained to identify a lymphocyte cell classification within a cell border and a non-lymphocyte cell classification within a cell border.

25. The method of claim 1, wherein the image-based biomarker prediction system is communicatively coupled to a pathology slide scanner system through a communication network, such that the image-based biomarker prediction system receives the digital image from the pathology slide scanner system over the communication network.

26. The method of claim 1, the method comprising:

applying, using the one or more processors, the plurality of tile images to the deep learning framework comprising one or more trained biomarker classification models, each trained biomarker classification model being trained to classify a different biomarker;

predicting, using the one or more processors, a biomarker classification for each of the plurality of tile images using the one or more trained biomarker classification models; and

generating a report containing the digital image and a digital overlay visualizing the predicted presence of the one or more biomarkers.

27. The method of claim 1, further comprising: identifying one or more therapies or lines of therapy associated with the predicted presence of the one or more biomarkers; and presenting the identified therapies or lines of therapy in a report.

28. The method of claim 1, further comprising: generating a report containing the digital image, a digital overlay visualizing the predicted presence of the one or more biomarkers, and eligible clinical trials.

29. A computing device configured to identifying biomarkers in digital image of a Hematoxylin and eosin (H&E) stained slide of target tissue, the computing device comprising:

one or more memories; and

one or more processors configured to,

receive the digital image;

perform an image tiling process on the digital image by separating the digital image into a plurality of tile images, where each of the plurality of tile images contains a different portion of the digital image;

apply the plurality of tile images to a multiscale deep learning framework comprising:

(i) one or more trained deep learning multiscale classifier models, each trained deep learning multiscale classifier models being trained to classify a different tissue classification for each tile image and to determine a tissue classification for each of the plurality of tile images, using the multiscale deep learning framework, and

(ii) a trained cell segmentation model that processes pixels of the digital image;

identify cells within the digital image using the trained cell segmentation model; and

from the tissue classification determined for each tile image and from the identified cells within the digital image, identify a predicted presence of one or more biomarkers associated with the digital image.

30. The computing device of claim 29, wherein the one or more processors are configured to:

perform the image tiling process on the digital image by applying a tiling mask to the digital image, to separate the digital image into the plurality of tile images.

31. The computing device of claim 29, wherein the one or more processors are further configured to:

receive a plurality of H&E slide training images from a training images dataset, each H&E slide training image having a label corresponding to a biomarker to be trained;

perform tile-based tissue classification analysis on each of the H&E slide training images;

perform a pixel-based cell segmentation analysis on each of the H&E slide training images;

optionally perform a tile-based biomarker classification analysis on each of the H&E slide training images; and

in response, generate the one or more trained deep learning multiscale classifier models.

32. The computing device of claim 31, wherein each H&E slide training image comprises a plurality of tile images each having a tile-level label.

33. The computing device of claim 31, wherein the one or more processors are further configured to:

for each H&E slide training image, perform a tile selection process that infers a class status for each tile image in the H&E slide training image;

based on inferred class status, discard tile images not corresponding to a desired class, before performing the tile-based tissue classification analysis on each of the H&E slide training images, such that the tile-based tissue classification analysis is performed on only selected tile images of the H&E slide training image.

34. The computing device of claim 29, wherein the one or more processors are further configured to:

train the one or more trained deep learning multiscale classifier models by:

receiving a molecular training dataset for a plurality of training tissue samples, the molecular training dataset comprising RNA transcriptome counts from sequencing of a substantially similar sample associated with each training tissue sample;

performing a clustering process on the molecular training dataset to identify one or more molecular data subsets each corresponding to a different biomarker;

for each of the one or more molecular data subsets identifying a plurality of digital images of H&E stained training slides of training tissue samples corresponding to the respective biomarker to an image-based biomarker prediction system having one or more processors;

performing tile-based tissue classification analysis on each of the H&E slide training images;

performing a pixel-based cell segmentation analysis on each of the H&E slide training images;

optionally performing a tile-based biomarker classification analysis on each of the H&E slide training images; and

in response, generating the one or more trained deep learning multiscale classifier models.

35. The computing device of claim 29, wherein the one or more biomarkers are selected from the group consisting of tumor-infiltrating lymphocytes (TILs), nucleus-to-cytyoplasm (NC) ratio, ploidy, signet ring morphology, and programmed death-ligand 1 (PD-L1).

36. The computing device of claim 29, wherein the one or more biomarkers is TILs, wherein the one or more processors are further configured to:

identify lymphocyte cells within the digital image using a trained lymphocyte segmentation model, by integrating the cell borders identified using the cell segmentation model with the lymphocytes identified using the lymphocyte segmentation model and by generating a nested classification of each cell.

37. The computing device of claim 29, wherein the one or more biomarkers is TILs, wherein the one or more processors are further configured to:

identify lymphocyte cells within the digital image using a trained lymphocyte segmentation model, by applying each of a plurality of tile images to a trained lymphocyte segmentation model and, for each tile image, by assigning a lymphocyte classification to one or more pixels within the tile image.

38. The computing device of claim 29, wherein the one or more biomarkers is TILs, wherein the one or more processors are further configured to:

identify, using the one or more processors, lymphocyte cells within the digital image using a trained lymphocyte segmentation model, wherein the lymphocyte segmentation model is a pixel-resolution two-dimensional UNet classification model trained to identify a lymphocyte cell classification within a cell border and a non-lymphocyte cell classification within a cell border.

39. The computing device of claim 29, wherein the computing device is communicatively coupled to a pathology slide scanner system through a communication network, such that the computing device receives the digital image from the pathology slide scanner system over the communication network.

40. The computing device of claim 29, wherein the computing device is contained within a pathology slide scanner system.

41. The computing device of claim 29, wherein at least one of the one or more processors of the computing device is contained within a pathology slide scanner system.

42. A computing device configured to identifying biomarkers in digital image of a Hematoxylin and eosin (H&E) stained slide of target tissue, the computing device comprising:

one or more memories; and

one or more processors configured to,

receive a molecular training dataset for a plurality of training tissue samples, the molecular training dataset comprising RNA transcriptome counts from sequencing of a substantially similar sample associated with each training tissue sample;

perform a clustering process on the molecular training dataset to identify one or more molecular data subsets each corresponding to a different biomarker;

for each of the one or more molecular data subsets, receive a plurality of digital images of H&E stained training slides of training tissue samples corresponding to the respective biomarker to an image-based biomarker prediction system having one or more processors;

generate, for each of the one or more molecular data subsets, a trained image-based biomarker classifier model, based on the plurality of digital images of the H&E stained training slides and a trained cell segmentation model that processes pixels of the digital image;

receive a subsequent digital image of an H&E stained slide of a subsequent tissue sample; and

apply the subsequent digital image to the trained image-based biomarker classifier model and the trained cell segmentation model to identify a predicted presence of one or more biomarkers for the subsequent tissue sample.