



(51) International Patent Classification:

G06V 10/82 (2022.01) G06N 3/0895 (2023.01)

G06V 20/69 (2022.01) G06V 10/25 (2022.01)

(21) International Application Number:

PCT/EP2023/060894

(22) International Filing Date:

26 April 2023 (26.04.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/337,771 03 May 2022 (03.05.2022) US

(71) Applicant: **BAYER AKTIENGESELLSCHAFT** [DE/DE]; Kaiser-Wilhelm-Allee 1, 51373 Leverkusen (DE).

(72) Inventors: **HÖHNE, Johannes**; c/o Bayer Aktiengesellschaft, Kaiser-Wilhelm-Allee 1, 51373 Leverkusen (DE). **CERSOVSKY, Josef**; c/o Bayer Aktiengesellschaft, Kaiser-Wilhelm-Allee 1, 51373 Leverkusen (DE). **LENGA, Matthias**; c/o Bayer Aktiengesellschaft, Kaiser-Wilhelm-Allee 1, 51373 Leverkusen (DE). **DE ZOETE, Jacob Coenraad**; c/o Bayer Aktiengesellschaft, Kaiser-Wilhelm-Allee 1, 51373 Leverkusen (DE). **SCHMITZ, Arndt**; c/o Bayer Aktiengesellschaft, Kaiser-Wilhelm-Allee 1, 51373 Leverkusen (DE). **BAL, Tricia**; c/o Bayer U.S. LLC, 100 Bayer Boulevard, Whippany, New Jersey 07981 (US). **PELEKANOU, Vasiliki**; c/o Bayer

U.S. LLC, 100 Bayer Boulevard, Whippany, New Jersey 07981 (US). **DI TOMASO, Emmanuelle**; c/o Bayer U.S. LLC, 100 Bayer Boulevard, Whippany, New Jersey 07981 (US).

(74) Agent: **BIP PATENTS**; c/o Bayer Intellectual Property GmbH, Alfred-Nobel-Straße 50, 40789 Monheim am Rhein (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: DYNAMIC SAMPLING STRATEGY FOR MULTIPLE-INSTANCE LEARNING

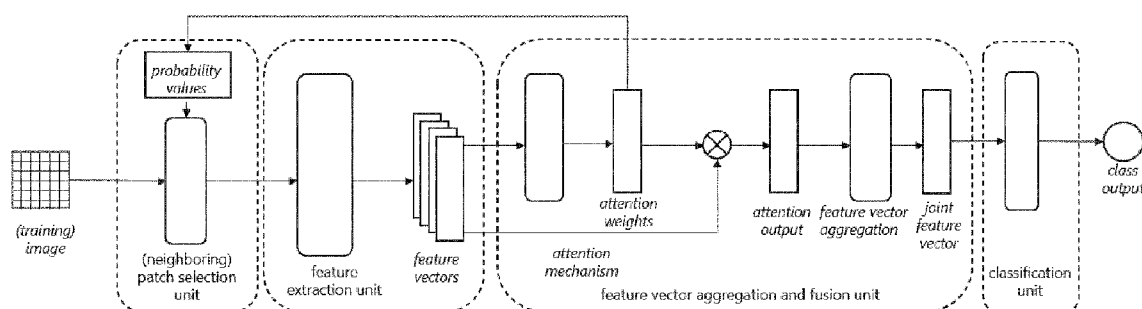


Fig. 4

(57) Abstract: Systems, methods, and computer programs disclosed herein relate to training a machine learning model and using the trained machine learning model to classify images, preferably medical images, using multiple-instance learning techniques. The machine learning model can be trained, and the trained machine learning model can be used for various purposes, in particular for the detection, identification and/or characterization of tumor types and/or gene mutations in tissues.



Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

- *with international search report (Art. 21(3))*

Dynamic sampling strategy for multiple-instance learning

FIELD

Systems, methods, and computer programs disclosed herein relate to training a machine learning model and using the trained machine learning model to classify images, preferably medical images, using multiple-instance learning techniques. The machine learning model can be trained, and the trained machine learning model can be used for various purposes, in particular for the detection, identification and/or characterization of tumor types and/or gene mutations in tissues.

BACKGROUND

Multi-instance learning is common for computer vision tasks, especially in medical image processing.

Multiple-instance learning uses training sets that consist of bags where each bag contains several instances that are either positive or negative examples for the class of interest, but only bag-level labels are given, and the instance-level labels are unknown during training.

In the field of image classification, a multi-instance learning approach can be applied when the images are very large, i.e., have a very large number of pixels. Instead of training a machine learning model based on the complete images, it can be trained based on patches.

A patch is a subregion of an image which is smaller than the original image.

From an image a number of patches can be generated, and a machine learning model can be trained to

- generate a patch embedding for each patch,
- aggregate patch embeddings into a bag-level representation, and
- classify the bag-level-representation into one of at least two classes.

Once the machine learning model is trained, it can be used to classify a new image into one of the two trained classes.

J. Hoehne *et al.* describe the use of such a multiple-instance learning approach to detect genetic alterations in tumor tissue samples: *Detecting genetic alterations in BRAF and NTRK as oncogenic drivers in digital pathology images: towards model generalization within and across multiple thyroid cohorts*, Proceedings of Machine Learning Research 156, 2021, pages 1-12.

When training the machine learning model, patches can be selected randomly. However, a random selection is inefficient because usually many patches are irrelevant for the classification result of an image. Thus, the task is to find a solution how to preferentially select such patches that have a higher relevance for the classification result.

SUMMARY

This task is solved by the independent patent claims. Preferred embodiments can be found in the dependent claims as well as in the description and the drawings.

Therefore, in a first aspect, the present disclosure provides a computer-implemented multiple-instance learning method for training a machine learning model for classifying images, the method comprising:

- (1) receiving training images, each training image being assigned to one of at least two classes,
- (2) for each training image:
 - (2.1) generating a plurality of patches and regions based on the training image, wherein the training image comprises the plurality of regions, and each region comprises at least one patch,
 - (2.2) assigning a probability value to each patch and/or each region,
 - (2.3) selecting a number of patches from the training image, each selected patch being selected according to the probability value of the patch or the region it comprises,

- (2.4) inputting the selected patches into the machine learning model, wherein the machine learning model is configured
- to generate a feature vector based on the selected patches and parameters of the machine learning model, and
 - to assign the feature vector to one of the at least two classes and thereby generating a classification result,
- (2.5) computing a loss based on a difference between the class to which the feature vector is assigned and the class to which the training image is assigned,
- (2.6) modifying parameters of the machine learning model based on the computed loss,
- (2.7) re-assigning probability values to patches and/or regions of the training image, wherein the probability values are determined based on a relevance of the patches and/or regions to the classification result,
- (2.8) repeating steps (2.3) to (2.8) several times until the loss reaches a defined minimum,
- (3) storing the trained machine learning model and/or using the trained machine learning model to classify one or more new images.

In another aspect, the present disclosure provides a computer system comprising:

a processor; and

a memory storing an application program configured to perform, when executed by the processor, an operation, the operation comprising:

- (1) receiving training images, each training image being assigned to one of at least two classes,
- (2) for each training image:
 - (2.1) generating a plurality of patches and regions based on the training image, wherein the training image comprises the plurality of regions, and each region comprises at least one patch,
 - (2.2) assigning a probability value to each patch and/or each region,
 - (2.3) selecting a number of patches from the training image, each selected patch being selected according to the probability value of the patch or the region it comprises,
 - (2.4) inputting the selected patches into the machine learning model, wherein the machine learning model is configured
 - to generate a feature vector based on the selected patches and parameters of the machine learning model, and
 - to assign the feature vector to one of the at least two classes and thereby generating a classification result,
 - (2.5) computing a loss based on a difference between the class to which the feature vector is assigned and the class to which the training image is assigned,
 - (2.6) modifying parameters of the machine learning model based on the computed loss,
 - (2.7) re-assigning probability values to patches and/or regions of the training image, wherein the probability values are determined based on a relevance of the patches and/or regions to the classification result,
 - (2.8) repeating steps (2.3) to (2.8) several times until the loss reaches a defined minimum,
- (3) storing the trained machine learning model and/or using the trained machine learning model to classify one or more new images.

In another aspect, the present disclosure provides a non-transitory computer readable medium having stored thereon software instructions that, when executed by a processor of a computer system, cause the computer system to execute the following steps:

- 5 (1) receiving training images, each training image being assigned to one of at least two classes,
- (2) for each training image:
 - (2.1) generating a plurality of patches and regions based on the training image, wherein the training image comprises the plurality of regions, and each region comprises at least one patch,
 - 10 (2.2) assigning a probability value to each patch and/or each region,
 - (2.3) selecting a number of patches from the training image, each selected patch being selected according to the probability value of the patch or the region it comprises,
 - (2.4) inputting the selected patches into the machine learning model, wherein the machine learning model is configured
 - 15 • to generate a feature vector based on the selected patches and parameters of the machine learning model, and
 - to assign the feature vector to one of the at least two classes and thereby generating a classification result,
 - (2.5) computing a loss based on a difference between the class to which the feature vector is assigned and the class to which the training image is assigned,
 - 20 (2.6) modifying parameters of the machine learning model based on the computed loss,
 - (2.7) re-assigning probability values to patches and/or regions of the training image, wherein the probability values are determined based on a relevance of the patches and/or regions to the classification result,
 - 25 (2.8) repeating steps (2.3) to (2.8) several times until the loss reaches a defined minimum,
- (3) storing the trained machine learning model and/or using the trained machine learning model to classify one or more new images.

30 In another aspect, the present disclosure relates to the use of a trained machine learning model for the detection, identification, and/or characterization of tumor types and/or gene mutations in tissues, wherein training of the machine learning model comprises:

- (1) receiving training images, each training image being assigned to one of at least two classes,
- (2) for each training image:
 - 35 (2.1) generating a plurality of patches and regions based on the training image, wherein the training image comprises the plurality of regions, and each region comprises at least one patch,
 - (2.2) assigning a probability value to each patch and/or each region,
 - (2.3) selecting a number of patches from the training image, each selected patch being selected according to the probability value of the patch or the region it comprises,
 - 40 (2.4) inputting the selected patches into the machine learning model, wherein the machine learning model is configured
 - to generate a feature vector based on the selected patches and parameters of the machine learning model, and

- to assign the feature vector to one of the at least two classes and thereby generating a classification result,
- (2.5) computing a loss based on a difference between the class to which the feature vector is assigned and the class to which the training image is assigned,
- 5 (2.6) modifying parameters of the machine learning model based on the computed loss,
- (2.7) re-assigning probability values to patches and/or regions of the training image, wherein the probability values are determined based on a relevance of the patches and/or regions to the classification result,
- (2.8) repeating steps (2.3) to (2.8) several times until the loss reaches a defined minimum,
- 10 (3) storing the trained machine learning model and/or using the trained machine learning model to classify one or more new images.

In another aspect, the present disclosure relates to a method for classifying an image, the method comprising:

- 15 (1) receiving training images, each training image being assigned to one of at least two classes,
- (2) for each training image:
 - (2.1) generating a plurality of patches and regions based on the training image, wherein the training image comprises the plurality of regions, and each region comprises at least one patch,
 - 20 (2.2) assigning a probability value to each patch and/or each region,
 - (2.3) selecting a number of patches from the training image, each selected patch being selected according to the probability value of the patch or the region it comprises,
 - (2.4) inputting the selected patches into the machine learning model, wherein the machine learning model is configured
 - 25 • to generate a feature vector based on the selected patches and parameters of the machine learning model, and
 - to assign the feature vector to one of the at least two classes and thereby generating a classification result,
 - 30 (2.5) computing a loss based on a difference between the class to which the feature vector is assigned and the class to which the training image is assigned,
 - (2.6) modifying parameters of the machine learning model based on the computed loss,
 - (2.7) re-assigning probability values to patches and/or regions of the training image, wherein the probability values are determined based on a relevance of the patches and/or regions to the classification result,
 - 35 (2.8) repeating steps (2.3) to (2.8) several times until the loss reaches a defined minimum, thereby obtaining a trained machine learning model,
- (3) receiving a new image,
- (4) generating a plurality of patches based on the new image,
- (5) inputting the patches into the trained machine learning model,
- 40 (6) receiving a classification result from the trained machine learning model,
- (7) outputting the classification result.

DETAILED DESCRIPTION

The invention will be more particularly elucidated below without distinguishing between the aspects of the disclosure (method, computer system, computer-readable storage medium, use). On the contrary, the following elucidations are intended to apply analogously to all the aspects of the disclosure, irrespective of in which context (method, computer system, computer-readable storage medium, use) they occur.

If steps are stated in an order in the present description or in the claims, this does not necessarily mean that the disclosure is restricted to the stated order. On the contrary, it is conceivable that the steps can also be executed in a different order or else in parallel to one another, unless one step builds upon another step, this absolutely requiring that the building step be executed subsequently (this being, however, clear in the individual case). The stated orders are thus preferred embodiments of the invention.

As used herein, the articles “a” and “an” are intended to include one or more items and may be used interchangeably with “one or more” and “at least one.” As used in the specification and the claims, the singular form of “a”, “an”, and “the” include plural referents, unless the context clearly dictates otherwise. Where only one item is intended, the term “one” or similar language is used. Also, as used herein, the terms “has”, “have”, “having”, or the like are intended to be open-ended terms. Further, the phrase “based on” is intended to mean “based at least partially on” unless explicitly stated otherwise. Further, the phrase “based on” may mean “in response to” and be indicative of a condition for automatically triggering a specified operation of an electronic device (e.g., a controller, a processor, a computing device, etc.) as appropriately referred to herein.

Some implementations of the present disclosure will be described more fully hereinafter with reference to the accompanying drawings, in which some, but not all implementations of the disclosure are shown. Indeed, various implementations of the disclosure may be embodied in many different forms and should not be construed as limited to the implementations set forth herein; rather, these example implementations are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the disclosure to those skilled in the art.

The present disclosure provides means for training a machine learning model and using the trained machine learning model for prediction purposes.

Such a “machine learning model”, as used herein, may be understood as a computer implemented data processing architecture. The machine learning model can receive input data and provide output data based on that input data and on parameters of the machine learning model. The machine learning model can learn a relation between input data and output data through training. In training, parameters of the machine learning model may be adjusted in order to provide a desired output for a given input.

The process of training a machine learning model involves providing a machine learning algorithm (that is the learning algorithm) with training data to learn from. The term “trained machine learning model” refers to the model artifact that is created by the training process. The training data must contain the correct answer, which is referred to as the target. The learning algorithm finds patterns in the training data that map input data to the target, and it outputs a trained machine learning model that captures these patterns.

In the training process, training data are inputted into the machine learning model and the machine learning model generates an output. The output is compared with the (known) target. Parameters of the machine learning model are modified in order to reduce the deviations between the output and the (known) target to a (defined) minimum.

In general, a loss function can be used for training, where the loss function can quantify the deviations between the output and the target. The loss function may be chosen in such a way that it rewards a wanted relation between output and target and/or penalizes an unwanted relation between an output and a target. Such a relation can be, e.g., a similarity, or a dissimilarity, or another relation.

A loss function can be used to calculate a loss for a given pair of output and target. The aim of the training process can be to modify (adjust) parameters of the machine learning model in order to reduce

the loss to a (defined) minimum. The loss can be, for example, a cross-entropy loss, in case of a binary classification it can be, for example a binary cross-entropy loss.

The machine learning model of the present disclosure is trained to assign an image to one of at least two classes.

- 5 The term “image” as used herein means a data structure that represents a spatial distribution of a physical signal. The spatial distribution may be of any dimension, for example 2D, 3D or 4D. The spatial distribution may be of any shape, for example forming a grid and thereby defining pixels, the grid being possibly irregular or regular. The physical signal may be any signal, for example color, level of gray, depth, surface or volume occupancy, such that the image may be a 2D or 3D RGB/grayscale/depth
10 image, or a 3D surface/volume occupancy model.

For simplicity, the invention is described herein mainly on the basis of two-dimensional images consisting of a rectangular array of pixels. However, this is not to be understood as limiting the invention to such images. Those skilled in machine learning based on image data will know how to apply the invention to image data comprising more dimensions and/or being in a different format.

- 15 In a preferred embodiment, the image is a medical image.

A “medical image” is a preferably visual representation of the human body or a part thereof or a visual representation of the body of an animal or a part thereof. Medical images can be used, e.g., for diagnostic and/or treatment purposes.

- 20 Techniques for generating medical images include X-ray radiography, computerized tomography, fluoroscopy, magnetic resonance imaging, ultrasonography, endoscopy, elastography, tactile imaging, thermography, microscopy, positron emission tomography and others.

Examples of medical images include CT (computer tomography) scans, X-ray images, MRI (magnetic resonance imaging) scans, fluorescein angiography images, OCT (optical coherence tomography) scans, histopathological images, ultrasound images.

- 25 In a preferred embodiment, the image is a whole slide histopathological image of a tissue of a human body.

In a preferred embodiment, the histopathological image is an image of a stained tissue sample. One or more dyes can be used to create the stained image. Usual dyes are hematoxylin and eosin.

The machine learning model is trained to assign the image to one of at least two classes.

- 30 The class may indicate whether the tissue shown in the image has a certain property or does not have the certain property.

The class may indicate whether there is a specific gene mutation in the tissue shown in the image.

The class may indicate whether the tissue depicted in the image is tumor tissue and/or may specify the type of tumor present in the tissue.

- 35 The class may indicate whether the subject from which the tissue depicted in the image originates has a particular disease or does not have the particular disease.

The class may indicate the severity of a particular disease.

Further options for classes are described below.

- 40 The training of the machine learning model of the present disclosure is performed based on training data. The training data comprise a plurality of training images. The term “plurality of training images” means at least 10, preferably at least 100 training images.

Each training image is labelled, i.e., it is assigned to one of the at least two classes. The labelling can be done, e.g., by one or more (medical) experts. For example, in histopathology, it is usually known what

the tissue is that is depicted in a histopathological image. It is possible, for example, to take a genetic analysis of a tissue sample and identify gene mutations. A slide and a slide image can then be generated from the tissue sample. The slide image can then be labeled accordingly with the information obtained from the gene analysis, for example.

5 Usually, the training images come from a number of different patients.

Especially with large images (e.g., whole slide images), usually not the complete images are used in the training and for prediction purposes, but only parts of them, due to the size of the images and the associated computing power required to process such images.

10 Therefore, for each training image, a plurality of patches and regions is generated from the training image.

A patch is usually one part of an image. The term “plurality of patches” means usually more than 100 patches, preferably more than 1000 patches. The number of patches and/or the shape of patches can be the same for each image or it can be different. Usually, the patches have a square or rectangular shape. In case of a quadratic or rectangular 2D image, the resolution of a patch is usually in the range of 32
15 pixels x 32 pixels to 10.000 pixels x 10.000 pixels, preferably in the range of 128 pixels x 128 pixel to 4096 pixels x 4096 pixels.

Fig. 1 shows schematically by way of example an image I which is divided into a plurality of patches P. In the example depicted in Fig. 1, the patches form a square grid in which each patch can be assigned an x-coordinate and a y-coordinate that determine its location within the grid. In Fig. 1, three of the
20 patches are marked with an arrow. Their x, y-coordinates are: (31, 30); (32, 31); (32, 32).

A “region” is a collection of patches within an image; in other words, a region is composed of a number of patches (at least one patch). The number of patches per region can be for example 10, 12, 15, 16, 20, 30, 50, 80, 100, 120, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200 or any other number. Preferably, a region includes at least 10 patches. Usually, but not necessarily, the regions are each
25 composed of the same number of patches. Usually, all regions have the same shape. Usually, but not necessarily, the regions are rectangular or square.

Fig. 2 shows schematically by way of example, an image I which is divided into $4 \times 4 = 16$ regions, each region comprising $8 \times 8 = 64$ patches. In Fig. 2, one patch P and one region R are highlighted by hatching.

30 In other word, a (training) image can be divided into a number of regions, and each region can be divided into a number of patches. In general, each patch is composed of a number of pixels or voxels; usually, the pixels/voxels are the smallest elements of an image.

It should be noted that the order in which regions and patches are created can be reversed. So, an image can be divided into regions first and the regions can then be divided into patches, or an image can be divided into patches and groups of patches can then be combined into regions.

35 Each patch and/or each region is assigned a probability value. The probability value of a patch determines the probability of selecting the patch. The probability value of a region determines the probability of selecting a patch that is located in the region.

At the beginning of the training, each patch and/or region can be assigned the same probability value. It is also possible that each patch and/or region is assigned a random probability value. It is also possible
40 that each patch and/or region is assigned a probability value based on a metric, wherein the metric may quantify, for example, the number of patches per region or a complexity of the region and/or patches in the region.

In the course of the training, the probability value is then adapted to the relevance of the patch and/or region for the classification result.

45 For classifying an image according to the present disclosure, a feature vector representing the image is generated. The feature vector serves as a basis for the classification.

During training of the machine learning model, each training image is subjected to a series of operations to generate the feature vector representing the training image.

In machine learning, a feature vector is an m -dimensional vector of numerical features that represent an object, wherein m is an integer greater than 0. Many algorithms in machine learning require a numerical representation of objects since such representations facilitate processing and statistical analysis. The term “feature vector” shall also include single values, matrices, tensors, and the like. The generation of a feature vector is often accompanied by a dimension reduction in order to reduce an object (e.g., an image) to those features that are important for the classification. Which features these are, the machine learning model learns during training.

- Examples of feature vector generation methods can be found in various textbooks and scientific publications (see, e.g., G.A. Tsihrintzis, L.C. Jain: *Machine Learning Paradigms: Advances in Deep Learning-based Technological Applications*, in: Learning and Analytics in Intelligent Systems Vol. 18, Springer Nature, 2020, ISBN: 9783030497248; K. Grzegorzczuk: *Vector representations of text data in deep learning*, Doctoral Dissertation, 2018, arXiv:1901.01695v1 [cs.CL]; M. Ilse *et al.*: Attention-based Deep Multiple-instance Learning, arXiv:1802.04712v4 [cs.LG]).

In the present case, the feature vector representing the (training) image is generated based on features of selected patches and optionally of neighbored patches and/or features of regions in which selected patches are located and/or aggregated features of further regions (regions not comprising selected patches).

- The machine learning model is trained based on training images. Each training image can be fed to the machine learning model one or more times; usually it is fed to the machine learning model multiple times. As described, usually not the entire image is fed to the model, but selected patches. A patch is not randomly selected but according to the probability value of the patch or the probability value of the region where the selected patch is located.
- For each selected patch, a patch embedding is generated based on the selected patches. A patch embedding is a feature vector representing a single patch.

- It is possible to additionally select one or more patches adjacent to each selected patch, generate patch embeddings representing such neighbored patches and fuse all patch embeddings into a feature vector (global embedding) representing the image and being used for classification, as described, e.g., in A. V. Konstantinov *et al.*: *Multi-Attention Multiple-instance Learning*, arXiv:2112.06071v1 [cs.LG].

A feature vector representing a (training) image is also referred to as “global embedding” in this disclosure.

- It is also possible to consider more distant neighbor patches and aggregate their patch embeddings into so-called regional embeddings representing adjacent regions. Patch embeddings of selected patches and regional embeddings are then aggregated into a global embedding which may serve as a basis for classification. This is described in US patent application No. 63/329,127, the content of which is incorporated herein by reference in its entirety.

- It is also possible to aggregate features of regions, where selected patches are located, into regional embeddings, which are then fused with the patch embeddings of selected patches and optionally adjacent patches into a global embedding. This is described in US patent application No. 63/334,903, the content of which is incorporated herein by reference in its entirety. The aforementioned patent application also describes that aggregated features of regions where no selected patches are located can also be included in the generation of a global embedding; this is also applicable to the present disclosure.

- Whenever a patch and/or a neighbor patch is selected, it is preferably selected according to the probability value of the patch or the probability value of the region comprising the respective patch.

Selected patches (and neighbored patches and/or patches of regions comprising selected patches and/or patches of adjacent regions, if applicable) are inputted into a feature extraction unit of the machine learning model which is configured to generate a patch embedding for each patch.

All patch embeddings are then aggregated into the global embedding representing the respective (training) image. For the aggregation, an attention mechanism can be used.

In machine learning, attention is a technique that mimics cognitive attention. The effect enhances some parts of the input data while diminishing other parts - the thought being that the machine learning model should devote more focus to that small but important part of the data. Learning which part of the data is more important than others is trained in the training phase. Details about attention mechanisms can be found, e.g., in: A. V. Konstantinov *et al.*: *Multi-Attention Multiple instance Learning*, arXiv:2112.06071v1 [cs.LG] and/or in M. Ilse *et al.*: *Attention-based Deep Multiple instance Learning*, arXiv:1802.04712v4 [cs.LG].

Attention weights are thus assigned to the features of the patches. These attention weights are based on learnable parameters of the machine learning model.

The feature vector (global embedding) representing the (training) image is then inputted into a classification unit. The classification unit is configured to assign the global embedding to one of at least two classes.

In a next step, it can be checked whether the classification result is correct. A loss can be computed using a loss function. The loss quantifies the difference between the class to which the global embedding is assigned and the class to which the training image is assigned.

Parameters of the machine learning model, including the parameters on which the attention weights are based, can be modified in order to reduce the loss (e.g., to a defined minimum). A gradient descent optimization procedure can be used to reduce the loss.

Before the next training cycle is started, probability values are reassigned to patches and/or regions. The probability values are determined based on the relevance of the patches and/or regions to the classification result. In other words, the probability values are determined based on the relevance of the assignment of the training image to the one of the two classes.

For example, relevance can be determined based on parameters of the machine learning model. Parameters of the machine learning model reflect the relevance of the patches and/or regions for the classification result.

During the training of the machine learning model, parameters of the machine learning model are modified to improve the prediction accuracy of the model. In other word, during training, the machine learning model learns which parameters need to be modified and in what way in order to achieve the most error-free mapping of the input data to the target data. If the probability values are derived from parameters of the machine learning model, the machine learning model learns at the same time which patches should be selected preferentially in order to make the training as efficient as possible.

In a preferred embodiment, attention weights are used to assign new (updated) probability values to the regions. For example, the re-assignment can be done in such a way that patches whose features have higher attention weights than other patches are assigned a higher probability value. For example, the re-assignment can be done in such a way that regions with patches whose features have higher attention weights than other regions are assigned a higher probability value.

For example, it is possible to sum the attention weights from all the patches of each region and then assign a probability value to the regions that is positively correlated with, for example proportional to, the sum of the attention weights. Other re-assignment options are also possible.

For example, it is also possible to assign a probability value to individual patches that is positively correlated with the attention weights of the patches' features, and to compute a probability value for a region by averaging the probability values of the patches that lie in the region (e.g., an arithmetic mean).

It is also possible to assign to a region the maximum probability value that a patch in the region has.

It is also possible to increase the probability value of such adjacent patches and/or regions that are adjacent to a patch and/or region that has/have a high probability value. The rationale behind this strategy is that patches in the immediate vicinity of relevant patches are also more likely to be relevant than patches further away.

The probability values can be updated after each training cycle or after a defined number of training cycles. Preferably, at the beginning of the training, the probability values are not changed after each training cycle. Preferably, a number of training cycles are run in which the probability values are kept constant / frozen. This allows the machine learning model to learn the relevance of individual patches, e.g., based on the attention weights before the attention weights influence the selection of patches.

After re-assigning probability values, patches are selected again. Patches are selected according to their (re-assigned) probability value, or according to the (re-assigned) probability value of the region in which they are located.

Patch embeddings (and optionally regional embeddings) are again generated and aggregated into a global embedding using the attention mechanism, and classification is again performed, and the classification result is again evaluated, and parameters of the model, including parameters on which the attention weights are based, are again modified to reduce the discrepancies between the class to which the global embedding is assigned and the class to which the training image is assigned.

The modified attention weights can then again be used to re-assign probability values and the next training cycle is initiated based on the re-assigned probability values until the machine learning model achieves a satisfactory classification result.

The process is repeated several times for many training images and preferably also several times for each training image, so that the machine learning model learns which features have a relevance for the classification result and weights them higher accordingly.

Fig. 3 shows schematically, by way of example, a preferred embodiment of the training of the machine learning model. The training comprises:

- (1) receiving training images, each training image being assigned to one of at least two classes,
- (2) for each training image:
 - (2.1) generating a plurality of patches and regions based on the training image, wherein the training image comprises the plurality of regions, and each region comprises at least one patch,
 - (2.2) assigning a probability value to each region,
 - (2.3) selecting a number of patches from the training image, each selected patch being selected according to the probability value of the region it comprises,
 - (2.4) inputting the selected patches into the machine learning model, wherein the machine learning model is configured
 - to generate a feature vector based on the selected patches using an attention mechanism, wherein the attention mechanism assigns attention weights to features of the selected patches, and
 - to assign the feature vector to one of the at least two classes,
 - (2.5) computing a loss based on a difference between the class to which the feature vector is assigned and the class to which the training image is assigned,
 - (2.6) modifying parameters of the machine learning model, including the attention weights, based on the computed loss,
 - (2.7) re-assigning probability values to regions of the training image, wherein regions that include patches whose features have higher attention weights are assigned higher probability values,
 - (2.8) repeating steps (2.3) to (2.8) several times until the loss reaches a defined minimum,
- (3) storing the trained machine learning model and/or using the trained machine learning model to classify one or more new images.

Fig. 4 shows schematically one example of a possible architecture of the machine learning model of the present disclosure. The machine learning model comprises a patch selection unit which is configured to select patches according to probability values. The machine learning model further comprises a feature extraction unit which is configured to generate patch embeddings representing single patches. The machine learning model further comprises a feature vector aggregation and fusion unit which is configured to generate a joint feature vector (global embedding) based on patch embeddings (and optionally regional embeddings). One or more attention mechanisms are used in the aggregation process, wherein such an attention mechanism assigns learnable attention weights to features of patches. During training, the attention weights are modified in order to improve the classification result. Modified attention weights are used to update the probability values. The machine learning model further comprises a classification unit (also referred to as classifier) which is configured to assign the global embedding to one of the at least two classes.

The term "unit" as used in this disclosure is not intended to imply that there is necessarily a separate unit performing the functions described. Rather, the term is intended to be understood to mean that computation means are present which perform the appropriate functions. These computation means are typically one or more processors configured to perform corresponding operations. Details are described below with reference to Fig. 5.

The trained machine learning model can be stored in a data storage, transmitted to another computer system, or used to classify one or more new images. The term "new" means that the corresponding image was not used during training.

The machine learning model can be trained to perform various tasks. Accordingly, a trained machine learning model can be used for various purposes.

In a preferred embodiment, the machine learning model of the present disclosure is trained and the trained machine learning model is used to detect, identify, and/or characterize tumor types and/or gene mutations in tissues.

The machine learning model can be trained and the trained machine learning model can be used to recognize a specific gene mutation and/or a specific tumor type, or to recognize multiple gene mutations and/or a specific tumor type.

The machine learning model can be trained and the trained machine learning model can be used to characterize the type or types of cancer a patient or subject has.

The machine learning model can be trained and the trained machine learning model can be used to select one or more effective therapies for the patient.

The machine learning model can be trained and the trained machine learning model can be used to determine how a patient is responding over time to a treatment and, if necessary, to select a new therapy or therapies for the patient as necessary. Correctly characterizing the type or types of cancer a patient has and, potentially, selecting one or more effective therapies for the patient can be crucial for the survival and overall wellbeing of that patient.

The machine learning model can be trained and the trained machine learning model can be used to determine whether a patient should be included or excluded from participating in a clinical trial.

The machine learning model can be trained and the trained machine learning model can be used to classify images of tumor tissue in one or more of the following classes: inflamed, non-inflamed, vascularized, non-vascularized, fibroblast-enriched, non-fibroblast-enriched (such classes are defined, e.g., in EP3639169A1).

The machine learning model can be trained and the trained machine learning model can be used to identify differentially expressed genes in a sample from a subject (e.g., a patient) having a cancer (e.g., a tumor).

The machine learning model can be trained and the trained machine learning model can be used to identify genes that are mutated in a sample from a subject having a cancer (e.g., a tumor).

The machine learning model can be trained and the trained machine learning model can be used to identify a cancer (e.g., a tumor) as a specific subtype of cancer selected.

Such uses may be useful for clinical purposes including, for example, selecting a treatment, monitoring cancer progression, assessing the efficacy of a treatment against a cancer, evaluating suitability of a patient for participating in a clinical trial, and/or determining a course of treatment for a subject (e.g., a patient).

The trained machine learning model may also be used for non-clinical purposes including (as a non-limiting example) research purposes such as, e.g., studying the mechanism of cancer development and/or biological pathways and/or biological processes involved in cancer, and developing new therapies for cancer based on such studies.

The machine learning model of the present disclosure is trained based on images and it generates predictions based on images. The images usually show the tissue of one or more subjects. The images can be created from tissue samples of a subject. The subject is usually a human, but may also be any mammal, including mice, rabbits, dogs, and monkeys.

The tissue sample may be any sample from a subject known or suspected of having cancerous cells or pre-cancerous cells.

The tissue sample may be from any source in the subject's body including, but not limited to, skin (including portions of the epidermis, dermis, and/or hypodermis), bone, bone marrow, brain, thymus, spleen, small intestine, appendix, colon, rectum, liver, gall bladder, pancreas, kidney, lung, ureter, bladder, urethra, uterus, ovary, cervix, scrotum, penis, prostate.

The tissue sample may be a piece of tissue, or some or all of an organ.

The tissue sample may be a cancerous tissue or organ or a tissue or organ suspected of having one or more cancerous cells.

The tissue sample may be from a healthy (e.g. non-cancerous) tissue or organ.

The tissue sample may include both healthy and cancerous cells and/or tissue.

In certain embodiments, one sample has been taken from a subject for analysis. In some embodiments, more than one (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more) samples may have been taken from a subject for analysis.

In some embodiments, one sample from a subject will be analyzed. In certain embodiments, more than one (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more) samples may be analyzed. If more than one sample from a subject is analyzed, the samples may have been procured at the same time (e.g., more than one sample may be taken in the same procedure), or the samples may have been taken at different times (e.g., during a different procedure including a procedure 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 days; 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 weeks; 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 months, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 years, or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 decades after a first procedure). A second or subsequent sample may be taken or obtained from the same region (e.g., from the same tumor or area of tissue) or a different region (including, e.g. a different tumor). A second or subsequent sample may be taken or obtained from the subject after one or more treatments and may be taken from the same region or a different region. As a non-limiting example, the second or subsequent sample may be useful in determining whether the cancer in each sample has different characteristics (e.g., in the case of samples taken from two physically separate tumors in a patient) or whether the cancer has responded to one or more treatments (e.g., in the case of two or more samples from the same tumor prior to and subsequent to a treatment).

Any of the samples described herein may have been obtained from a subject using any known technique. In some embodiments, the sample may have been obtained from a surgical procedure (e.g., laparoscopic surgery, microscopically controlled surgery, or endoscopy), bone marrow biopsy, punch biopsy, endoscopic biopsy, or needle biopsy (e.g., a fine-needle aspiration, core needle biopsy, vacuum-assisted biopsy, or image-guided biopsy).

Detection, identification, and/or characterization of tumor types may be applied to any cancer and any tumor. Exemplary cancers include, but are not limited to, adrenocortical carcinoma, bladder urothelial carcinoma, breast invasive carcinoma, cervical squamous cell carcinoma, endocervical adenocarcinoma, colon adenocarcinoma, esophageal carcinoma, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, ovarian serous cystadenocarcinoma, pancreatic adenocarcinoma, prostate adenocarcinoma, rectal adenocarcinoma, skin cutaneous melanoma, stomach adenocarcinoma, thyroid carcinoma, uterine corpus endometrial carcinoma, and cholangiocarcinoma.

The machine learning model can be trained and the trained machine learning model can be used to detect, identify and/or characterize gene mutations in tissue samples.

Examples of genes related to proliferation of cancer or response rates of molecular target drugs include HER2, TOP2A, HER3, EGFR, P53, and MET. Examples of tyrosine kinase related genes include ALK, FLT3, AXL, FLT4 (VEGFR3), DDR1, FMS(CSF1R), DDR2, EGFR(ERBB1), HER4(ERBB4), EML4-ALK, IGF1R, EPHA1, INSR, EPHA2, IRR(INSRR), EPHA3, KIT, EPHA4, LTK, EPHA5, MER(MERTK), EPHA6, MET, EPHA7, MUSK, EPHA8, NPM1-ALK, EPHB1, PDGFR α (PDGFRA), EPHB2, PDGFR β (PDGFRB), EPHB3, RET, EPHB4, RON(MST1R), FGFR1, ROS(ROS1), FGFR2, TIE2(TEK), FGFR3, TRKA(NTRK1), FGFR4, TRKB(NTRK2), FLT1(VEGFR1), and TRKC(NTRK3). Examples of breast cancer related genes include ATM, BRCA1, BRCA2, BRCA3, CCND1, E-Cadherin, ERBB2, ETV6, FGFR1, HRAS, KRAS, NRAS, NTRK3, p53, and PTEN. Examples of genes related to carcinoid tumors include BCL2, BRD4, CCND1, CDKN1A, CDKN2A, CTNNB1, HES1, MAP2, MEN1, NF1, NOTCH1, NUT, RAF, SDHD, and VEGFA. Examples of colorectal cancer related genes include APC, MSH6, AXIN2, MYH, BMPR1A, p53, DCC, PMS2, KRAS2 (or Ki-ras), PTEN, MLH1, SMAD4, MSH2, STK11, and MSH6. Examples of lung cancer related genes include ALK, PTEN, CCND1, RASSF1A, CDKN2A, RB1, EGFR, RET, EML4, ROS1, KRAS2, TP53, and MYC. Examples of liver cancer related genes include Axin1, MALAT1, b-catenin, p16 INK4A, c-ERBB-2, p53, CTNNB1, RB1, Cyclin D1, SMAD2, EGFR, SMAD4, IGFR2, TCF1, and KRAS. Examples of kidney cancer related genes include Alpha, PRCC, ASPSCR1, PSF, CLTC, TFE3, p54nrb/NONO, and TFEB. Examples of thyroid cancer related genes include AKAP10, NTRK1, AKAP9, RET, BRAF, TFG, ELE1, TPM3, H4/D10S170, and TPR. Examples of ovarian cancer related genes include AKT2, MDM2, BCL2, MYC, BRCA1, NCOA4, CDKN2A, p53, ERBB2, PIK3CA, GATA4, RB, HRAS, RET, KRAS, and RNASET2. Examples of prostate cancer related genes include AR, KLK3, BRCA2, MYC, CDKN1B, NKX3.1, EZH2, p53, GSTP1, and PTEN. Examples of bone tumor related genes include CDH11, COL12A1, CNBP, OMD, COL1A1, THRAP3, COL4A5, and USP6.

In a preferred embodiment, the machine learning model is trained and used for classification of tissue types on the basis of whole slide images. Preferably, the machine learning model is trained and used for identification of gene mutations, such as BRAF mutations and/or NTRK fusions, as described in WO2020229152A1 and/or J. Hoehne *et al.*: *Detecting genetic alterations in BRAF and NTRK as oncogenic drivers in digital pathology images: towards model generalization within and across multiple thyroid cohorts*, Proceedings of Machine Learning Research 156, 2021, pages 1-12, the contents of which are incorporated by reference in their entirety into this specification.

For example, the machine learning model can be trained to detect signs of the presence of oncogenic drivers in patient tissue images stained with hematoxylin and eosin.

F. Penault-Llorca *et al.* describe a testing algorithm for identification of patients with TRK fusion cancer (see J. Clin. Pathol., 2019, 72, 460-467). The algorithm comprises immunohistochemistry (IHC) studies, fluorescence in situ hybridization (FISH) and next-generation sequencing.

Immunohistochemistry provides a routine method to detect protein expression of NTRK genes. However, performing immunohistochemistry requires additional tissue section(s) and time to proceed and interpret (following hematoxylin and eosin initial staining based on which tumor diagnosis is performed), skills and the correlation between protein expression and gene fusion status is not trivial.

Interpretation of IHC results requires the skills of a trained and certified medical professional pathologist.

Similar practical challenges hold true for other molecular assays such as FISH.

5 Next-generation sequencing provides a precise method to detect NTRK gene fusions. However, performing gene analyses for each patient is expensive, tissue consuming (not always feasible when available tissue specimen is minimal, as in diagnostic biopsies), not universally available in various geographic locations or diagnostic laboratories/healthcare institutions and, due to the low incidence of NTRK oncogenic fusions, inefficient.

10 There is therefore a need for a comparatively rapid and inexpensive method to detect signs of the presence of specific tumors.

It is proposed to train a machine learning model as described in this disclosure to assign histopathological images of tissues from patients to one of at least two classes, where one class comprises images showing tissue in which a specific gene mutation is present, such as NTRK or BRAF.

15 It is proposed to use the trained machine learning model as a preliminary test. Patients in whom the specific mutation can be detected are then subjected to a standard examination such as IHC, FISH and/or next-generation sequencing to verify the finding.

Additional studies may also be considered, such as other forms of medical imaging (CT scans, MRI, etc.) that can be co-assessed using AI to generate multimodal biomarkers/characteristics for diagnostic purposes.

20 The machine *learning* model of the present disclosure can, e.g., be used to

- a) detect NTRK fusion events in one or more indications,
- b) detect NTRK fusion events in other indications than in those being trained on (i.e., an algorithm trained on thyroid data sets is useful in lung cancer data sets),
- c) detect NTRK fusion events involving other TRK family members (i.e., an algorithm trained on 25 NTRK1, NTRK3 fusions is useful to predict also NTRK2 fusions),
- d) detect NTRK fusion events involving other fusion partners (i.e., an algorithm trained on LMNA-fusion data sets is useful also in TPM3-fusion data sets),
- e) discover novel fusion partners (i.e., an algorithm trained on known fusion events might predict a fusion in a new data set which is then confirmed via molecular assay to involve a not yet 30 described fusion partner of a NTRK family member),
- f) catalyze the diagnostic workflow and clinical management of patients offering a rapid, tissue-sparing, low-cost method to indicate the presence of NTRK-fusions (and ultimately others) and identifying patients that merit further downstream molecular profiling so as to provide precision medicines targeting specific molecular aberrations (e.g. NTRK-fusion inhibitors),
- 35 g) identify specific genetic aberrations based on histological specimen can additionally be used to confirm/exclude or re-label certain tumor diagnosis, in cases the presence or absence of this/these alterations(s) is pathognomonic of specific tumors.

40 Identification of specific genetic aberrations based on histological specimen can additionally be used to confirm/exclude or re-label certain tumor diagnosis, in cases the presence or absence of this/these alterations(s) is pathognomonic of specific tumors.

Histopathological images used for training and prediction of the machine learning model can be obtained from patients by biopsy or surgical resection specimens.

In a preferred embodiment, a histopathological image is a microscopic image of tumor tissue of a human patient. The magnification factor is preferably in the range of 10 to 60, more preferably in the range of

20 to 40, whereas a magnification factor of, e.g., "20" means that a distance of 0.05 mm in the tumor tissue corresponds to a distance of 1 mm in the image (0.05 mm x 20 = 1 mm).

In a preferred embodiment, the histopathological image is a whole-slide image.

In a preferred embodiment, the histopathological image is an image of a stained tumor tissue sample.

5 One or more dyes can be used to create the stained images. Preferred dyes are hematoxylin and eosin.

Methods for creating histopathological images, in particular stained whole-slide microscopy images, are extensively described in scientific literature and textbooks (see e.g. S. K. Suvana *et al.*: *Bancroft's Theory and Practice of Histological Techniques*, 8th Ed., Elsevier 2019, ISBN 978-0-7020-6864-5; A. F. Frangi *et al.*: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, 21st International Conference Granada, Spain, 2018 Proceedings, Part II, ISBN 978-030-00933-5; L.C. Junqueira *et al.*: *Histologie*, Springer 2001, ISBN: 978-354-041858-0; N. Coudray *et al.*: *Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning*, Nature Medicine, Vol. 24, 2018, pages 1559-1567).

15 The machine learning model can also be configured to generate a probability value, the probability value indicating the probability of a patient suffering from cancer, e.g., caused by an NTRK oncogenic fusion. The probability value can be outputted to a user and/or stored in a database. The probability value can be a real number in the range from 0 to 1, whereas a probability value of 0 usually means that it is impossible that the cancer is caused by an NTRK oncogenic fusion, and a probability value of 1 usually means that there is no doubt that the cancer is caused by an NTRK oncogenic fusion. The probability value can also be expressed by a percentage.

20 In a preferred embodiment of the present invention, the probability value is compared with a predefined threshold value. In the event the probability value is lower than the threshold value, the probability that the patient suffers from cancer caused by an NTRK oncogenic fusion is low; treating the patient with a Trk inhibitor is not indicated; further investigations are required in order to determine the cause of cancer. In the event the probability value equals the threshold value or is greater than the threshold value, it is reasonable to assume that the cancer is caused by an NTRK oncogenic fusion; the treatment of the patient with a Trk inhibitor can be indicated; further investigations to verify the assumption can be initiated (e.g., performing a genetic analysis of the tumor tissue).

25 The threshold value can be a value between 0.5 and 0.9999999999, e.g. 0.8 (80%) or 0.81 (81 %) or 0.82 (82%) or 0.83 (83%) or 0.84 (84%) or 0.85 (85%) or 0.86 (86%) or 0.87 (87 %) or 0.88 (88%) or 0.89 (89%) or 0.9 (90%) or 0.91 (91 %) or 0.92 (92%) or 0.93 (93%) or 0.94 (94%) or 0.95 (95%) or 0.96 (96%) or 0.97 (97 %) or 0.98 (98%) or 0.99 (99%) or any other value (percentage). The threshold value can be determined by a medical expert.

30 Besides a histopathological image, additional patient data can also be included in the classification. Additional patient data can be, e.g., anatomic or physiology data of the patient, such as information about patient's height and weight, gender, age, vital parameters (such as blood pressure, breathing frequency and heart rate), tumor grades, ICD-9 classification, oxygenation of tumor, degree of metastasis of tumor, blood count value tumor indicator value like PA value, information about the tissue the histopathological image is created from (e.g. tissue type, organ), further symptoms, medical history etc. Also, the pathology report of the histopathological images can be used for classification, using text mining approaches. Also, a next generation sequencing raw data set which does not cover the TRK genes' sequences can be used for classification.

35 The operations in accordance with the teachings herein may be performed by at least one computer specially constructed for the desired purposes or general purpose computer specially configured for the desired purpose by at least one computer program stored in a typically non-transitory computer readable storage medium.

The term “non-transitory” is used herein to exclude transitory, propagating signals or waves, but to otherwise include any volatile or non-volatile computer memory technology suitable to the application.

5 The term “computer” should be broadly construed to cover any kind of electronic device with data processing capabilities, including, by way of non-limiting example, personal computers, servers, embedded cores, computing system, communication devices, processors (e.g., digital signal processor (DSP)), microcontrollers, field programmable gate array (FPGA), application specific integrated circuit (ASIC), etc.) and other electronic computing devices.

10 The term “process” as used above is intended to include any type of computation or manipulation or transformation of data represented as physical, e.g., electronic, phenomena which may occur or reside e.g., within registers and/or memories of at least one computer or processor. The term processor includes a single processing unit or a plurality of distributed or remote such units.

Fig. 5 illustrates a computer system (1) according to some example implementations of the present disclosure in more detail. The computer may include one or more of each of a number of components such as, for example, processing unit (20) connected to a memory (50) (e.g., storage device).

15 The processing unit (20) may be composed of one or more processors alone or in combination with one or more memories. The processing unit is generally any piece of computer hardware that is capable of processing information such as, for example, data, computer programs and/or other suitable electronic information. The processing unit is composed of a collection of electronic circuits some of which may be packaged as an integrated circuit or multiple interconnected integrated circuits (an integrated circuit at times more commonly referred to as a “chip”). The processing unit may be configured to execute
20 computer programs, which may be stored onboard the processing unit or otherwise stored in the memory (50) of the same or another computer.

The processing unit (20) may be a number of processors, a multi-core processor or some other type of processor, depending on the particular implementation. Further, the processing unit may be implemented
25 using a number of heterogeneous processor systems in which a main processor is present with one or more secondary processors on a single chip. As another illustrative example, the processing unit may be a symmetric multi-processor system containing multiple processors of the same type. In yet another example, the processing unit may be embodied as or otherwise include one or more ASICs, FPGAs or the like. Thus, although the processing unit may be capable of executing a computer program to perform
30 one or more functions, the processing unit of various examples may be capable of performing one or more functions without the aid of a computer program. In either instance, the processing unit may be appropriately programmed to perform functions or operations according to example implementations of the present disclosure.

35 The memory (50) is generally any piece of computer hardware that is capable of storing information such as, for example, data, computer programs (e.g., computer-readable program code (60)) and/or other suitable information either on a temporary basis and/or a permanent basis. The memory may include volatile and/or non-volatile memory, and may be fixed or removable. Examples of suitable memory include random access memory (RAM), read-only memory (ROM), a hard drive, a flash memory, a thumb drive, a removable computer diskette, an optical disk, a magnetic tape or some combination of
40 the above. Optical disks may include compact disk – read only memory (CD-ROM), compact disk – read/write (CD-R/W), DVD, Blu-ray disk or the like. In various instances, the memory may be referred to as a computer-readable storage medium. The computer-readable storage medium is a non-transitory device capable of storing information, and is distinguishable from computer-readable transmission media such as electronic transitory signals capable of carrying information from one location to another.
45 Computer-readable medium as described herein may generally refer to a computer-readable storage medium or computer-readable transmission medium.

In addition to the memory (50), the processing unit (20) may also be connected to one or more interfaces for displaying, transmitting and/or receiving information. The interfaces may include one or more

communications interfaces and/or one or more user interfaces. The communications interface(s) may be configured to transmit and/or receive information, such as to and/or from other computer(s), network(s), database(s) or the like. The communications interface may be configured to transmit and/or receive information by physical (wired) and/or wireless communications links. The communications interface(s) may include interface(s) (41) to connect to a network, such as using technologies such as cellular telephone, Wi-Fi, satellite, cable, digital subscriber line (DSL), fiber optics and the like. In some examples, the communications interface(s) may include one or more short-range communications interfaces (42) configured to connect devices using short-range communications technologies such as NFC, RFID, Bluetooth, Bluetooth LE, ZigBee, infrared (e.g., IrDA) or the like.

The user interfaces may include a display (30). The display may be configured to present or otherwise display information to a user, suitable examples of which include a liquid crystal display (LCD), light-emitting diode display (LED), plasma display panel (PDP) or the like. The user input interface(s) (11) may be wired or wireless, and may be configured to receive information from a user into the computer system (1), such as for processing, storage and/or display. Suitable examples of user input interfaces include a microphone, image or video capture device, keyboard or keypad, joystick, touch-sensitive surface (separate from or integrated into a touchscreen) or the like. In some examples, the user interfaces may include automatic identification and data capture (AIDC) technology (12) for machine-readable information. This may include barcode, radio frequency identification (RFID), magnetic stripes, optical character recognition (OCR), integrated circuit card (ICC), and the like. The user interfaces may further include one or more interfaces for communicating with peripherals such as printers and the like.

As indicated above, program code instructions may be stored in memory, and executed by processing unit that is thereby programmed, to implement functions of the systems, subsystems, tools and their respective elements described herein. As will be appreciated, any suitable program code instructions may be loaded onto a computer or other programmable apparatus from a computer-readable storage medium to produce a particular machine, such that the particular machine becomes a means for implementing the functions specified herein. These program code instructions may also be stored in a computer-readable storage medium that can direct a computer, processing unit or other programmable apparatus to function in a particular manner to thereby generate a particular machine or particular article of manufacture. The instructions stored in the computer-readable storage medium may produce an article of manufacture, where the article of manufacture becomes a means for implementing functions described herein. The program code instructions may be retrieved from a computer-readable storage medium and loaded into a computer, processing unit or other programmable apparatus to configure the computer, processing unit or other programmable apparatus to execute operations to be performed on or by the computer, processing unit or other programmable apparatus.

Retrieval, loading and execution of the program code instructions may be performed sequentially such that one instruction is retrieved, loaded and executed at a time. In some example implementations, retrieval, loading and/or execution may be performed in parallel such that multiple instructions are retrieved, loaded, and/or executed together. Execution of the program code instructions may produce a computer-implemented process such that the instructions executed by the computer, processing circuitry or other programmable apparatus provide operations for implementing functions described herein.

Execution of instructions by processing unit, or storage of instructions in a computer-readable storage medium, supports combinations of operations for performing the specified functions. In this manner, a computer system (1) may include processing unit (20) and a computer-readable storage medium or memory (50) coupled to the processing circuitry, where the processing circuitry is configured to execute computer-readable program code (60) stored in the memory. It will also be understood that one or more functions, and combinations of functions, may be implemented by special purpose hardware-based computer systems and/or processing circuitry which perform the specified functions, or combinations of special purpose hardware and program code instructions.

CLAIMS

1. A computer-implemented method comprising:

(1) receiving training images, each training image being assigned to one of at least two classes,

(2) for each training image:

- 5 (2.1) generating a plurality of patches and regions based on the training image, wherein the training image comprises the plurality of regions, and each region comprises at least one patch,
- (2.2) assigning a probability value to each patch and/or each region,
- 10 (2.3) selecting a number of patches from the training image, each selected patch being selected according to the probability value of the patch or the region it comprises,
- (2.4) inputting the selected patches into a machine learning model, wherein the machine learning model is configured
- to generate a feature vector based on the selected patches and parameters of the machine learning model, and
 - 15 ▪ to assign the feature vector to one of the at least two classes and thereby generating a classification result,
- (2.5) computing a loss based on a difference between the class to which the feature vector is assigned and the class to which the training image is assigned,
- (2.6) modifying parameters of the machine learning model based on the computed loss,
- 20 (2.7) re-assigning probability values to patches and/or regions of the training image, wherein the probability values are determined based on a relevance of the patches and/or regions to the classification result,
- (2.8) repeating steps (2.3) to (2.8) several times until the loss reaches a defined minimum,
- 25 (3) storing the trained machine learning model and/or using the trained machine learning model to classify one or more new images.

2. The method according to claim 1, wherein relevance of the patches and/or regions is determined based on parameters of the machine learning model.

30 3. The method according to claim 1 or 2, wherein an attention mechanism is used to generate the feature vector based on the selected patches, wherein the attention mechanism assigns learnable attention weights to features of the selected patches.

35 4. The method according to claim 3, wherein patches whose features have higher attention weights are re-assigned higher probability values and/or regions with patches whose features have higher attention weights are assigned higher probability values.

40 5. The method according to any one of claims 3 to 5, wherein re-assigned probability values are determined by summing the attention weights from all the patches of each region and then assigning a probability value to the regions that is positively correlated with the sum of the attention weights.

6. The method according to any one of claims 3 to 5, wherein a probability value is re-assigned to each patch, wherein the probability value is positively correlated with the attention weights of the patches' features.

5 7. The method according to claim 6, wherein a probability value for each region is determined based on the probability values of the patches comprises by the region.

8. The method according to claim 7, wherein the probability value of the region is equal to the probability value of the patch in the region with the highest probability value.

10

9. The method according to claim 7, wherein the probability value of the region is the arithmetic mean of the probability values of the patches in the region.

15

10. The method according to any one of claims 1 to 9, wherein the probability value of such neighboring patches and/or regions that are adjacent to a patch and/or region that has/have a high probability value is increased during re-assigning.

20

11. The method according to any one of claims 1 to 10, wherein one class of the at least two classes comprises images showing tissue in which a specific gene mutation is present, preferably a mutation affecting one or more of the following genes: HER2, TOP2A, HER3, EGFR, P53, MET, ALK, FLT3, AXL, FLT4, DDR2, EGFR, HER4, EML4-ALK, IGF1R, EPHA1, INSR, EPHA2, IRR, EPHA3, KIT, EPHA4, LTK, EPHA5, MER, EPHA6, MET, EPHA7, MUSK, EPHA8, NPM1-ALK, EPHB1, PDGFR α , EPHB2, PDGFR β , EPHB3, RET, EPHB4, RON, FGFR1, ROS, FGFR2, TIE2, FGFR3, TRKA, FGFR4, TRKB, FLT1, TRKC, ATM, BRCA1, BRCA2, BRCA3, CCND1, E-Cadherin, ERBB2, ETV6, FGFR1, HRAS, KRAS, NRAS, NTRK3, p53, PTEN, BCL2, BRD4, CCND1, CDKN1A, CDKN2A, CTNNB1, HES1, MAP2, MEN1, NF1, NOTCH1, NUT, RAF, SDHD, VEGFA, APC, MSH6, AXIN2, MYH, BMPR1A, p53, DCC, PMS2, KRAS2, PTEN, MLH1, SMAD4, MSH2, STK11, MSH6, PTEN, CCND1, RASSF1A, CDKN2A, RB1, EGFR, RET, EML4, ROS1, KRAS2, TP53, MYC, Axin1, MALAT1, b-catenin, p16 INK4A, c-ERBB-2, p53, CTNNB1, RB1, Cyclin D1, SMAD2, EGFR, SMAD4, IGFR2, TCF1, KRAS, Alpha, PRCC, ASPSCR1, PSF, CLTC, TFE3, p54nrb/NONO, TFEB, AKAP10, NTRK1, AKAP9, RET, BRAF, TFG, ELE1, TPM3, H4/D10S170, TPR, AKT2, MDM2, BCL2, MYC, BRCA1, NCOA4, CDKN2A, p53, ERBB2, PIK3CA, GATA4, RB, HRAS, RET, KRAS, RNASET2, AR, KLK3, BRCA2, MYC, CDKN1B, NKX3.1, EZH2, p53, GSTP1, CDH11, COL12A1, CNBP, OMD, COL1A1, THRAP3, COL4A5, USP6.

35

12. The method according to any one of claims 1 to 11, wherein each training image is a medical image, preferably a whole slide image, most preferably a histopathological image of tissue from a patient stained with hematoxylin and eosin.

40

13. The method according to any one of claims 1 to 12, further comprising:

- (4) receiving a new image,
- (5) generating a plurality of patches based on the new image,
- (6) inputting the patches into the trained machine learning model,
- (7) receiving a classification result from the trained machine learning model,
- (8) outputting the classification result.

45

14. The method according to any one of claims 1 to 13, wherein the machine learning model is trained, and the trained machine learning model is used to assign histopathological images of tissues from patients to one of at least two classes, wherein one class comprises images showing tissue in which a NTRK or BRAF gene mutation is present.

5

15. A computer system comprising:

a processor; and

a memory storing an application program configured to perform, when executed by the processor, an operation, the operation comprising:

- 10 (1) receiving training images, each training image being assigned to one of at least two classes,
- (2) for each training image:
- (2.1) generating a plurality of patches and regions based on the training image, wherein the training image comprises the plurality of regions, and each region comprises at least one patch,
- 15 (2.2) assigning a probability value to each patch and/or each region,
- (2.3) selecting a number of patches from the training image, each selected patch being selected according to the probability value of the patch or the region it comprises,
- (2.4) inputting the selected patches into the machine learning model, wherein the machine learning model is configured
- 20 • to generate a feature vector based on the selected patches and parameters of the machine learning model, and
- to assign the feature vector to one of the at least two classes and thereby generating a classification result,
- (2.5) computing a loss based on a difference between the class to which the feature vector is assigned and the class to which the training image is assigned,
- 25 (2.6) modifying parameters of the machine learning model based on the computed loss,
- (2.7) re-assigning probability values to patches and/or regions of the training image, wherein the probability values are determined based on a relevance of the patches and/or regions to the classification result,
- 30 (2.8) repeating steps (2.3) to (2.8) several times until the loss reaches a defined minimum,
- (3) storing the trained machine learning model and/or using the trained machine learning model to classify one or more new images.

16. A non-transitory computer readable medium having stored thereon software instructions that, when executed by a processor of a computer system, cause the computer system to execute the following steps:

- 35 (1) receiving training images, each training image being assigned to one of at least two classes,
- (2) for each training image:
- (2.1) generating a plurality of patches and regions based on the training image, wherein the training image comprises the plurality of regions, and each region comprises at least one patch,
- 40 (2.2) assigning a probability value to each patch and/or each region,

- (2.3) selecting a number of patches from the training image, each selected patch being selected according to the probability value of the patch or the region it comprises,
- (2.4) inputting the selected patches into the machine learning model, wherein the machine learning model is configured
- 5 • to generate a feature vector based on the selected patches and parameters of the machine learning model, and
- to assign the feature vector to one of the at least two classes and thereby generating a classification result,
- (2.5) computing a loss based on a difference between the class to which the feature vector is assigned and the class to which the training image is assigned,
- 10 (2.6) modifying parameters of the machine learning model based on the computed loss,
- (2.7) re-assigning probability values to patches and/or regions of the training image, wherein the probability values are determined based on a relevance of the patches and/or regions to the classification result,
- 15 (2.8) repeating steps (2.3) to (2.8) several times until the loss reaches a defined minimum,
- (3) storing the trained machine learning model and/or using the trained machine learning model to classify one or more new images.

17. Use of a trained machine learning model for the detection, identification, and/or characterization of tumor types and/or gene mutations in tissues, wherein training of the machine learning model comprises:

- (1) receiving training images, each training image being assigned to one of at least two classes,
- (2) for each training image:
- (2.9) generating a plurality of patches and regions based on the training image, wherein the training image comprises the plurality of regions, and each region comprises at least one patch,
- 25 (2.10) assigning a probability value to each patch and/or each region,
- (2.11) selecting a number of patches from the training image, each selected patch being selected according to the probability value of the patch or the region it comprises,
- (2.12) inputting the selected patches into the machine learning model, wherein the machine learning model is configured
- 30 • to generate a feature vector based on the selected patches and parameters of the machine learning model, and
- to assign the feature vector to one of the at least two classes and thereby generating a classification result,
- 35 (2.13) computing a loss based on a difference between the class to which the feature vector is assigned and the class to which the training image is assigned,
- (2.14) modifying parameters of the machine learning model based on the computed loss,
- (2.15) re-assigning probability values to patches and/or regions of the training image, wherein the probability values are determined based on a relevance of the patches and/or regions to the classification result,
- 40 (2.16) repeating steps (2.3) to (2.8) several times until the loss reaches a defined minimum,
- (3) storing the trained machine learning model and/or using the trained machine learning model to classify one or more new images.

Figures

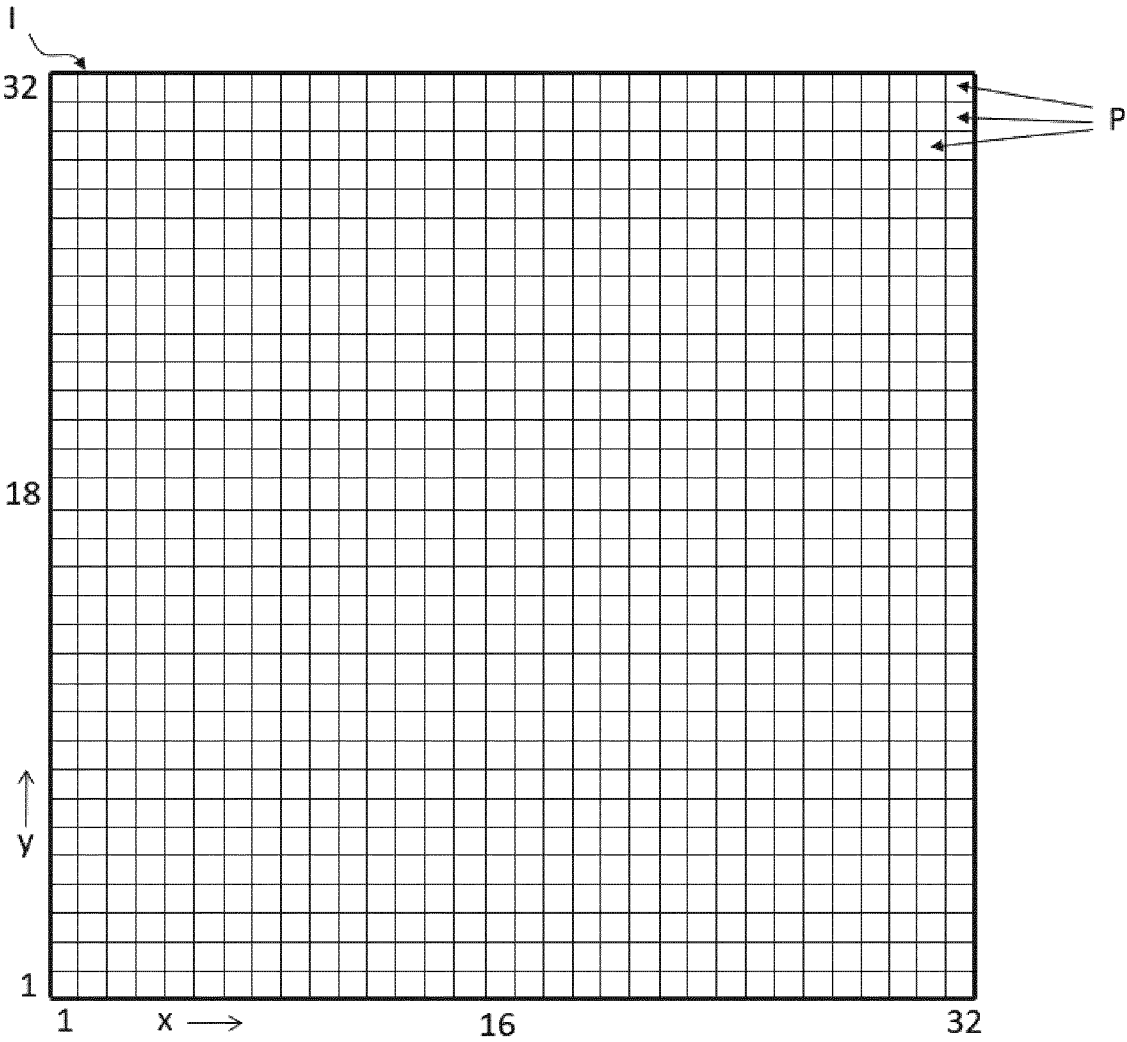


Fig. 1

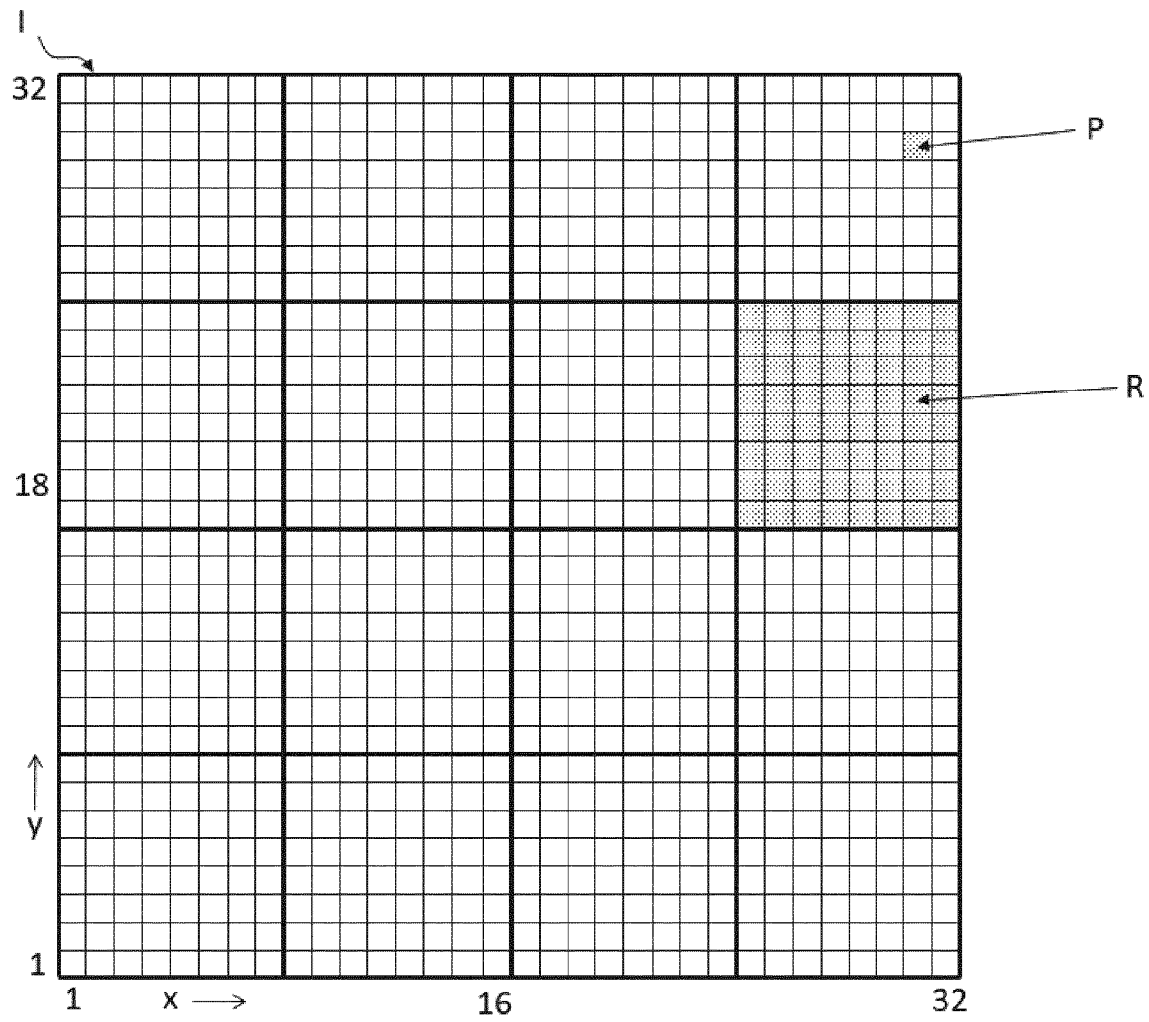


Fig. 2

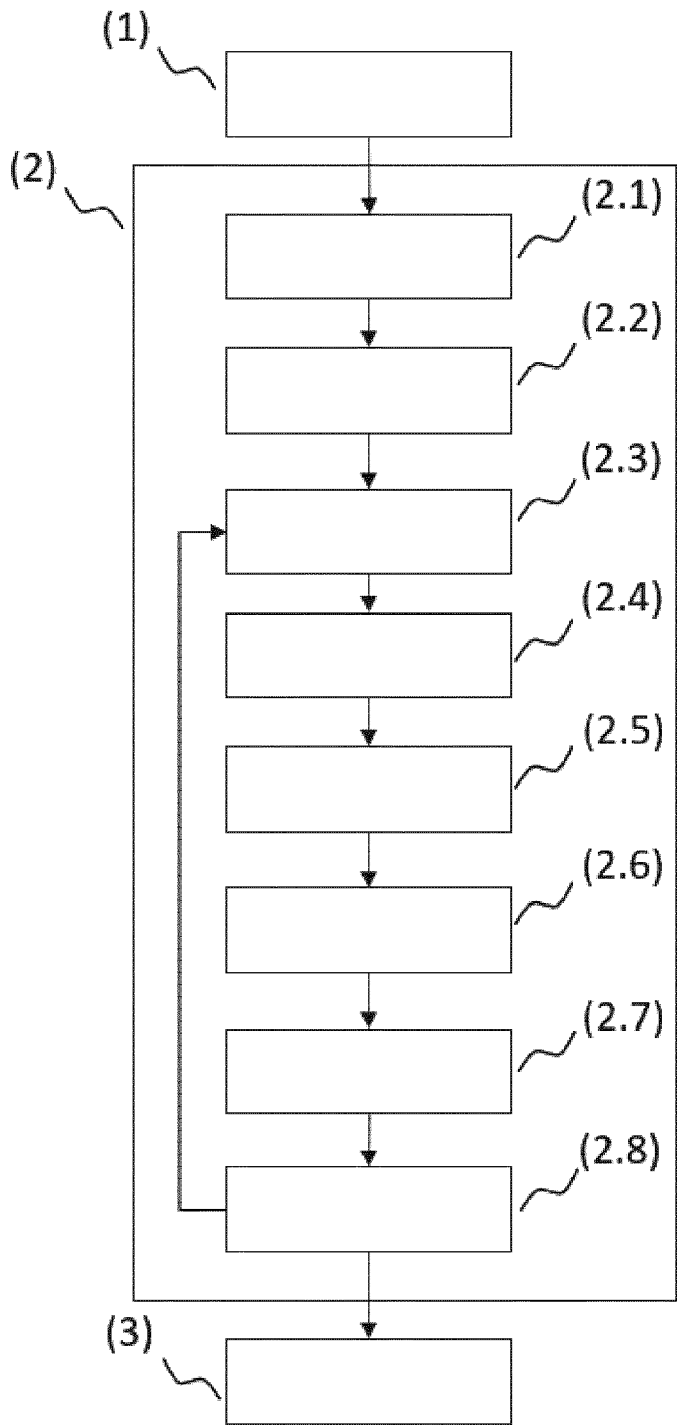


Fig. 3

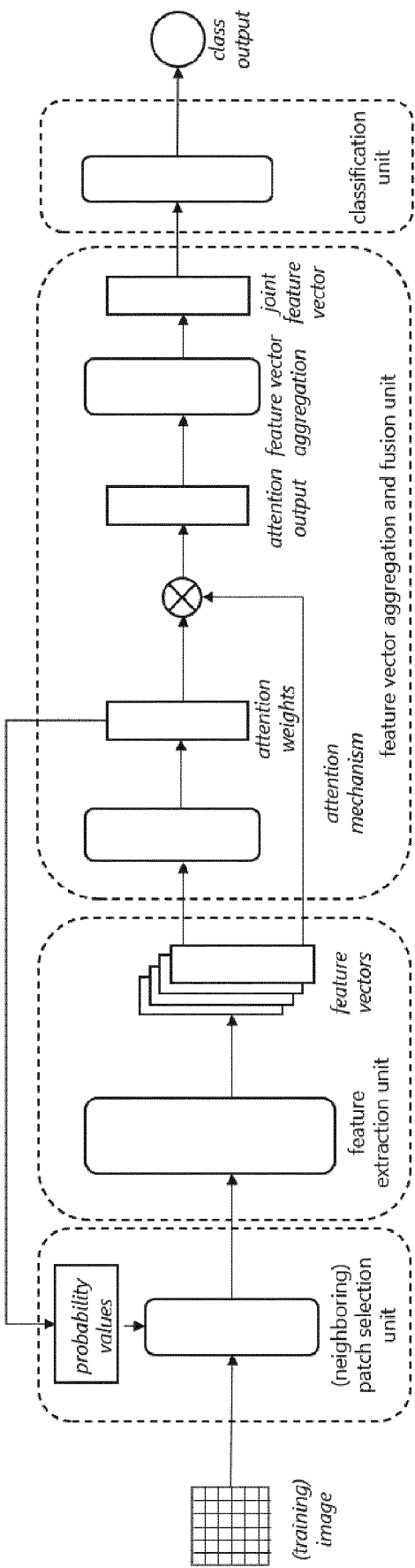


Fig. 4

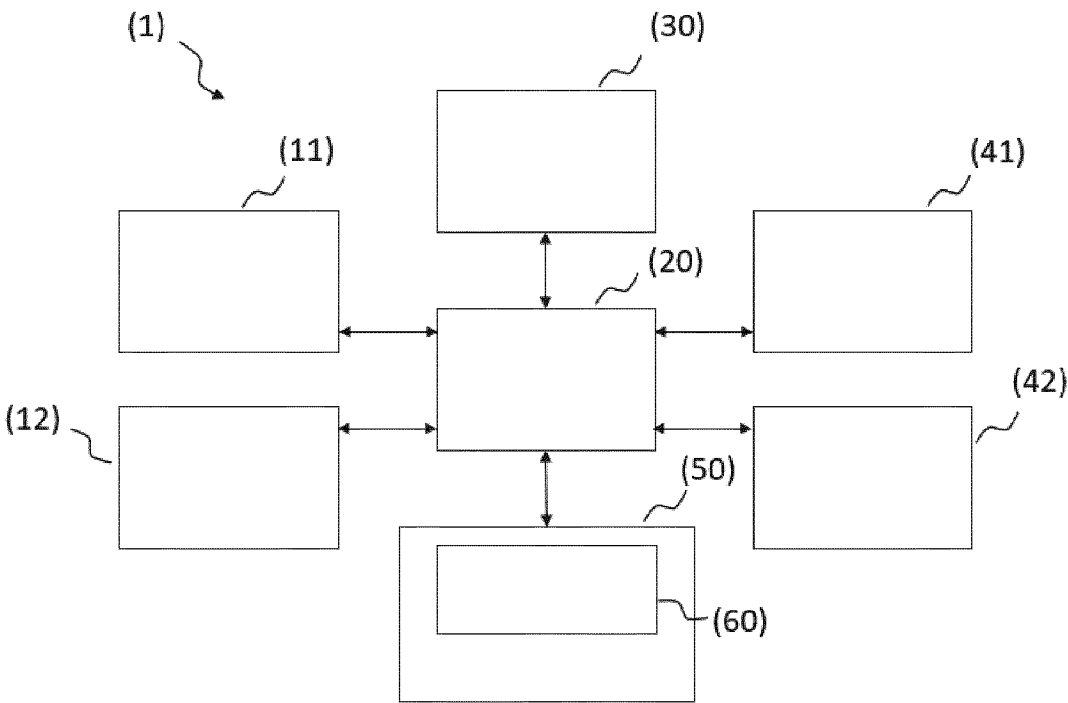


Fig. 5

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2023/060894

A. CLASSIFICATION OF SUBJECT MATTER INV. G06V10/82 G06V20/69 G06N3/0895 G06V10/25 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06V G06N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TARKHAN ALIASGHAR ET AL: "Attention-Based Deep Multiple Instance Learning with Adaptive Instance Sampling", 2022 IEEE 19TH INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING (ISBI), 28 March 2022 (2022-03-28), XP034116877, the whole document -----	1-17
X	SHEN YIQING ET AL: "Sampling Based Tumor Recognition in Whole-Slide Histology Image With Deep Learning Approaches", IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, vol. 19, no. 4, 25 February 2021 (2021-02-25), pages 2431-2441, XP011916783, sections 2 and 3 ----- --/--	1, 2, 10-17
☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
18 July 2023		26/07/2023
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Bouganis, Alexandros

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2023/060894

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ANDREI V KONSTANTINOV ET AL: "Multi-Attention Multiple Instance Learning", ARXIV.ORG, 11 December 2021 (2021-12-11), XP091115525, Retrieved from the Internet: URL:https://arxiv.org/pdf/2112.06071.pdf> the whole document -----	1-17
A	HÖHNE JOHANNES ET AL: "Detecting genetic alterations in BRAF and NTRK as oncogenic drivers in digital pathology images: towards model generalization within and across multiple thyroid cohorts", PROCEEDINGS OF THE MICCAI WORKSHOP ON COMPUTATIONAL PATHOLOGY, vol. 156, 27 September 2021 (2021-09-27), XP093064231, abstract sections 3 and 4 -----	1-17