Package 'PathoFMPred'

November 12, 2026

Type Package

Title Tumour Feature Models from Pathology Representations

Version 0.1.0

URL https://github.com/tkcaccia/PathoFMPred

BugReports https://github.com/tkcaccia/PathoFMPred/issues

Description Builds and applies portable cancer-specific partial least squares (PLS) regression and partial least squares with linear discriminant analysis (PLS-LDA) model collections from pathology foundation-model representations. Feature and outcome tables are linked by explicit patient identifiers, repeated slide rows are pooled at patient level, and model development uses nested cross-validation before final fitting. The source package contains no fitted foundation-model parameters, embeddings, patient-level data, or atlas registry. Users may download a separately licensed, checksum-verified Prov-GigaPath collection after installation, or load authorized collections from local files. Outputs are research estimates without external-validation or clinical-use claims. The methods build on Wold et al. (2001) <doi:10.1016/S0169-7439(01)00155-1> and Barker and Rayens (2003) <doi:10.1002/cem.785>.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 4.6.0)

Imports digest, fastPLS (>= 0.4), ggplot2, rmarkdown, stats, utils

Suggests knitr, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Config/roxygen2/version 8.1.0

NeedsCompilation no

Author Aamilah Ismail [aut],
Martin Ocharo [aut],
Giulia Canarutto [aut],
Moussa Kassim Mohamed [aut],

Dalia Ahmed [aut],
Brendon Price [aut],
Dupe Ojo [aut],
Ekene Emmanuel Nweke [aut],
Alessia Vignoli [aut],
Silvano Piazza [aut],
Dinesh Gupta [aut],
Leonardo Tenori [aut],
Stefano Cacciatore [aut, cre]

Maintainer Stefano Cacciatore <tkcaccia@gmail.com>

Repository CRAN

Date/Publication 2026-11-12 11:20:08 UTC

Contents

available_foundation_models 2
available_pathofm_models 3
compare_pathofm_models 4
create_pathofmpred_object 4
example_pathofm_features 6
fetch_pathofmpred_models 6
pathofmpred_control 7
pathofm_model_registry 9
pathofm_sample_report 9
plot_pathofm_binary 10
plot_pathofm_radar 11
plot_titan_binary 11
plot_titan_radar 12
predict_pathofm 12
predict_pathofmpred_object 13
predict_titan 14
validate_pathofmpred_object 15

Index **16**

`available_foundation_models` *List locally available foundation-model representations*

Description

List locally available foundation-model representations

Usage

```
available_foundation_models()
```

Value

A character vector. Availability means that at least one fitted research model collection is installed; it does not imply external validation or permission to redistribute the collection.

`available_pathofm_models`

List available cancer-specific models

Description

List available cancer-specific models

Usage

```
available_pathofm_models(
  foundation_model = NULL,
  cancer = NULL,
  outcome_type = NULL,
  family = NULL,
  include_limited_evidence = FALSE
)

available_titan_models(
  cancer = NULL,
  outcome_type = NULL,
  family = NULL,
  include_limited_evidence = FALSE
)
```

Arguments

`foundation_model`	One or more of 'TITAN', 'GigaSSL', or 'ProvGigaPath'.
`cancer`	Optional TCGA cancer code or vector of codes.
`outcome_type`	Optional 'continuous' or 'binary' filter.
`family`	Optional endpoint-family filter.
`include_limited_evidence`	Include limited-sample models only when they crossed the representation-specific matched effect threshold. Models tested below the threshold never enter prediction or reporting. The default is 'FALSE'.

Value

A filtered model-registry data frame.

`compare_pathofm_models` *Compare representation-specific models for one cancer-endpoint task*

Description

Lists fitted reference objects for the same task while distinguishing object availability from representation-specific matched-atlas effect-threshold crossing. The inventory is a partial operationalization of the multi-representation atlas. The function does not automatically select the largest internal estimate or combine representation-specific predictions.

Usage

```
compare_pathofm_models(cancer, endpoint, outcome_type = NULL, family = NULL,
  include_limited_evidence = FALSE)
```

Arguments

`cancer`	TCGA cancer code.
`endpoint`	Exact endpoint label.
`outcome_type`	Optional continuous or binary filter.
`family`	Optional endpoint-family filter.
`include_limited_evidence`	Include limited-sample models only when they crossed the representation-specific matched effect threshold.

Value

A registry data frame with object-availability, matched-threshold, resource-scope and selection-guidance fields.

`create_pathofmpred_object` *Build a portable PathoFMPred model collection*

Description

Feature and outcome tables are linked only through ‘id_column‘. Outcome IDs must be unique. Feature IDs may repeat, for example when a patient has more than one slide, and are pooled before model development.

Usage

```
create_pathofmpred_object(
  feature_table,
  outcome_table,
  id_column,
  aggregation = c("mean", "median"),
  foundation_model = "custom",
  feature_columns = NULL,
  outcome_columns = NULL,
  outcome_types = NULL,
  positive_class = NULL,
  control = pathofmpred_control(),
  output_file = NULL
)
```

Arguments

feature_table	Data frame, matrix, or CSV path containing identifiers and numeric foundation-model features.
outcome_table	Data frame or CSV path containing one unique row per identifier and one or more outcomes.
id_column	Name of the identifier column present in both tables.
aggregation	How repeated feature rows are pooled: ‘mean‘ or ‘median‘.
foundation_model	Label stored in the resulting object.
feature_columns	Optional numeric feature columns. By default all columns except ‘id_column‘ are used.
outcome_columns	Optional outcomes to fit. By default all columns except ‘id_column‘ are considered.
outcome_types	Optional named character vector containing ‘continuous‘ or ‘binary‘. Otherwise numeric two-level outcomes are binary and other numeric outcomes are continuous.
positive_class	Event-class label for binary outcomes. Supply either one value when fitting a single binary outcome or a named character vector keyed by endpoint. Zero/one and FALSE/TRUE outcomes default to one and TRUE; all other binary labels require an explicit value.
control	Settings created by [pathofmpred_control()].
output_file	Optional ‘.rds‘ destination.

Value

A ‘PathoFMPredObject‘ containing fitted models, a registry, nested-CV summaries, and a linkage/missingness audit.

example_pathofm_features

Construct a synthetic feature vector for examples and smoke tests

Description

The vector is the across-model mean of stored training-feature means for the requested cancer and is not a patient record.

Usage

```
example_pathofm_features(
  cancer = "COAD",
  foundation_model = "ProvGigaPath",
  include_limited_evidence = FALSE
)

example_titan_features(cancer = "BRCA", include_limited_evidence = FALSE)
```

Arguments

cancer	A cancer code with locally installed models.
foundation_model	One of 'TITAN', 'GigaSSL', or 'ProvGigaPath'.
include_limited_evidence	Use limited-evidence models when constructing the synthetic across-model feature mean. The default is 'FALSE'.

Value

A one-row data frame containing the named features required by the selected representation.

fetch_pathofmpred_models

Download user-authorized PathoFMPred model collections

Description

The source package contains no fitted models. For 'ProvGigaPath', the default is a versioned GitHub release asset whose SHA-256 digest is pinned in the package. It is downloaded only when this function is called. Other representations require an explicit authorized URL and SHA-256 digest. TITAN-derived objects are not downloadable through this helper.

Usage

```
fetch_pathofmpred_models(
  foundation_model = "ProvGigaPath",
  destination = NULL,
  base_url = NULL,
  expected_sha256 = NULL,
  overwrite = FALSE,
  quiet = FALSE
)
```

Arguments

foundation_model	'ProvGigaPath' or 'GigaSSL'.
destination	Explicit directory in which model collections will be stored. No directory is created unless the function is called with this argument.
base_url	Optional HTTPS release directory, or a 'file://' URL for an authorized local mirror. The pinned Prov-GigaPath release is the default when this is omitted.
expected_sha256	Optional named character vector containing a 64-character SHA-256 digest for every requested representation. The pinned Prov-GigaPath digest is the default for its release.
overwrite	Replace an existing file.
quiet	Passed to [utils::download.file()].

Value

Named paths to the downloaded model collections.

---

pathofmpred_control *Define model-development settings*

Description

Define model-development settings

Usage

```
pathofmpred_control(
  components = 1:20,
  outer_folds = 5L,
  inner_folds = 5L,
  repeats = 1L,
  scaling = "centering",
  method = "plssvd",
```

```
  backend = NULL,
  binary_selection = "AUROC",
  continuous_selection = "auto",
  seed = 1L,
  minimum_continuous = 50L,
  minimum_binary_class = 20L,
  rsvd_oversample = NULL,
  rsvd_power = NULL
)
```

Arguments

components	Candidate PLS component counts.
outer_folds, inner_folds	Numbers of nested cross-validation folds.
repeats	Number of independently seeded nested evaluations.
scaling, method, backend	Arguments passed to fastPLS.
binary_selection	Inner-CV objective for binary outcomes. AUROC is the default because the resulting score ranking is independent of prevalence- specific class priors.
continuous_selection	Inner-CV objective for continuous outcomes.
seed	Initial random seed.
minimum_continuous	Minimum labelled patients for a continuous model.
minimum_binary_class	Minimum patients in each binary class.
rsvd_oversample, rsvd_power	Optional overrides. 'NULL' preserves the installed fastPLS defaults.

Value

A settings list for [create_pathofmpred_object()].

Examples

```
settings <- pathofmpred_control(
  components = 1:3, outer_folds = 3, inner_folds = 3, seed = 42
)
settings$components
```

pathofm_model_registry

Read registries from locally available model collections

Description

The public package contains no fitted-model registry. This function reads metadata from collections that the user has installed under the applicable access terms. An empty data frame is returned when none is installed.

Usage

```
pathofm_model_registry()

titan_model_registry()
```

Value

A data frame with one row per locally available fitted model.

pathofm_sample_report *Create a single-sample pathology-foundation-model research output*

Description

Create a single-sample pathology-foundation-model research output

Usage

```
pathofm_sample_report(
  cancer,
  features,
  foundation_model = c("TITAN", "GigaSSL", "ProvGigaPath"),
  patient_id = NULL,
  output_file = NULL,
  format = c("html", "pdf", "both"),
  clinical_context = NULL,
  include_limited_evidence = FALSE,
  quiet = TRUE
)

titan_sample_report(
  cancer,
  features,
  patient_id = NULL,
```

```
  output_file = NULL,
  format = c("html", "pdf", "both"),
  clinical_context = NULL,
  include_limited_evidence = FALSE,
  quiet = TRUE
)
```

Arguments

cancer	A single TCGA cancer code.
features	One or more slide-feature rows for one patient.
foundation_model	One of 'TITAN', 'GigaSSL', or 'ProvGigaPath'.
patient_id	Optional patient/sample identifier for all rows.
output_file	Explicit file stem or '.html'/'.pdf' path. Required; no report is written unless the caller chooses a destination.
format	One of 'html', 'pdf', or 'both'.
clinical_context	Optional named list with 'shared_features', 'pathology', and 'source' character fields. This information is displayed as supplied and is not used by any prediction model.
include_limited_evidence	Explicitly opt in to limited-sample models that still crossed the matched effect threshold for the selected cancer and representation. Tested-below-threshold objects never enter reports.
quiet	Passed to [rmarkdown::render()].

Value

Invisibly, the generated file path(s).

---

plot_pathofm_binary *Plot PathoFMPred binary outputs*

Description

Representation-neutral alias for [plot_titan_binary()]. Binary score ranks remain explicitly uncalibrated ranks, not probabilities.

Usage

```
plot_pathofm_binary(predictions)
```

Arguments

predictions	Output from [predict_pathofm()] for one patient.

Value

A ggplot object.

`plot_pathofm_radar` *Radar plot of PathoFMPred continuous outputs*

Description

Representation-neutral alias for [plot_titan_radar()]. The selected foundation model is carried in the prediction rows returned by [predict_pathofm()]; no representation is selected inside the plotting function.

Usage

```
plot_pathofm_radar(predictions, max_endpoints = 12L)
```

Arguments

`predictions`	Output from [predict_pathofm()] for one patient.
`max_endpoints`	Maximum number of endpoints, prioritising extremes.

Value

A ggplot object.

`plot_titan_binary` *Plot binary PLS-LDA outputs*

Description

Plot binary PLS-LDA outputs

Usage

```
plot_titan_binary(predictions)
```

Arguments

`predictions`	Output from [predict_pathofm()] for one patient.

Value

A ggplot object. Scores are displayed as TCGA out-of-fold score ranks explicitly labelled "not probability".

plot_titan_radar *Radar plot of continuous prediction percentiles*

Description

Radar plot of continuous prediction percentiles

Usage

```
plot_titan_radar(predictions, max_endpoints = 12L)
```

Arguments

predictions	Output from [predict_pathofm()] for one patient.
max_endpoints	Maximum number of endpoints, prioritising extremes.

Value

A ggplot object. Endpoint names and original model predictions are placed at the outer radar corners to avoid overlap with the profile.

predict_pathofm *Apply supported models for one pathology foundation-model representation*

Description

Multiple rows with the same ‘patient_id‘ are interpreted as multiple slides and mean-pooled before prediction, matching model development.

Usage

```
predict_pathofm(
  cancer,
  features,
  foundation_model = c("TITAN", "GigaSSL", "ProvGigaPath"),
  patient_id = NULL,
  outcome_type = c("continuous", "binary"),
  family = NULL,
  warn_ood = TRUE,
  include_limited_evidence = FALSE
)
```

Arguments

cancer	One TCGA cancer code per row, or one code recycled to all rows.
features	A matrix or data frame containing the correctly named features for 'foundation_model'.
foundation_model	One of 'TITAN', 'GigaSSL', or 'ProvGigaPath'.
patient_id	Optional identifier per row. Repeated identifiers are pooled.
outcome_type	One or more of 'continuous' and 'binary'.
family	Optional endpoint-family filter.
warn_ood	Warn when over 5 percent of dimensions fall outside training ranges.
include_limited_evidence	Explicitly opt in to limited-sample models that still crossed the representation-specific matched effect threshold. Models tested below the threshold never enter prediction or reporting.

Value

A long data frame with one row per patient and applicable model. For atlas collections, fields prefixed 'matched_tss_' describe the representation-specific sensitivity analysis that held complete TCGA tissue-source-site codes apart. 'repeated_auc' and 'repeated_pr_auc' are missing when five-repeat validation was not performed for that model; they are never filled with a primary-screen estimate. These internal performance fields do not validate an individual prediction.

---

predict_pathofmpred_object

Apply a user-created PathoFMPred object

Description

This function applies an object returned by [create_pathofmpred_object()]. It is the supported inference route for locally constructed TITAN objects and for future foundation-model representations that are not distributed by the package. Repeated feature rows are pooled using the aggregation rule recorded when the object was created, unless 'aggregation' is supplied.

Usage

```
predict_pathofmpred_object(
  object,
  features,
  patient_id = NULL,
  id_column = NULL,
  outcome_type = c("continuous", "binary"),
  endpoints = NULL,
  aggregation = NULL,
  warn_ood = TRUE
)
```

Arguments

object	A ‘PathoFMPredObject‘ or the path to an ‘.rds‘ file containing one.
features	A data frame or matrix containing the named features used to fit ‘object‘. An identifier column may be selected with ‘id_column‘.
patient_id	Optional identifier per feature row. Use either ‘patient_id‘ or ‘id_column‘, not both. Repeated identifiers are pooled.
id_column	Optional identifier-column name in ‘features‘.
outcome_type	One or both of ‘continuous‘ and ‘binary‘.
endpoints	Optional endpoint names to apply.
aggregation	Optional ‘mean‘ or ‘median‘ override. By default the object’s model-development aggregation is used.
warn_ood	Warn when more than 5 percent of dimensions are outside the fitted training range.

Value

A long data frame with one row per patient and fitted endpoint. Binary outputs include the original event-class label. Scores are uncalibrated and are not probabilities.

---

predict_titan *Apply the legacy TITAN model interface*

Description

Backward-compatible wrapper for [predict_pathofm()] with ‘foundation_model = "TITAN"‘.

Usage

```
predict_titan(
  cancer,
  features,
  patient_id = NULL,
  outcome_type = c("continuous", "binary"),
  family = NULL,
  warn_ood = TRUE,
  include_limited_evidence = FALSE
)
```

Arguments

cancer	One TCGA cancer code per row, or one code recycled to all rows.
features	A matrix or data frame containing the correctly named features for ‘foundation_model‘.
patient_id	Optional identifier per row. Repeated identifiers are pooled.

`outcome_type` One or more of ‘continuous‘ and ‘binary‘.

`family` Optional endpoint-family filter.

`warn_ood` Warn when over 5 percent of dimensions fall outside training ranges.

`include_limited_evidence`
Explicitly opt in to limited-sample models that still crossed the representation-specific matched effect threshold. Models tested below the threshold never enter prediction or reporting.

Value

A ‘pathofm_predictions‘ data frame with one row for each patient and available TITAN endpoint. Continuous rows contain predictions in the recorded outcome scale. Binary rows contain an uncalibrated discriminant score and a class call; the score is not a probability.

`validate_pathofmpred_object`
Validate a PathoFMPred model collection

Description

Validate a PathoFMPred model collection

Usage

```
validate_pathofmpred_object(object)
```

Arguments

`object` An object created by [create_pathofmpred_object()] or a distributed PathoFM-Pred model collection.

Value

‘TRUE‘, invisibly, when the object is valid.

Index

available_foundation_models, 2
available_pathofm_models, 3
available_titan_models
(available_pathofm_models), 3

compare_pathofm_models, 4
create_pathofmpred_object, 4

example_pathofm_features, 6
example_titan_features
(example_pathofm_features), 6

fetch_pathofmpred_models, 6

pathofm_model_registry, 9
pathofm_sample_report, 9
pathofmpred_control, 7
plot_pathofm_binary, 10
plot_pathofm_radar, 11
plot_titan_binary, 11
plot_titan_radar, 12
predict_pathofm, 12
predict_pathofmpred_object, 13
predict_titan, 14

titan_model_registry
(pathofm_model_registry), 9
titan_sample_report
(pathofm_sample_report), 9

validate_pathofmpred_object, 15