Universidade Federal do Pará

Instituto de Ciências Exatas e Naturais

Faculdade de Computação

Eduardo Gil S. Cardoso

Analyzing the Impact of Dimensionality Reduction Over Human Intestinal Absorption Prediction Through Machine Learning Algorithms

Belém, PA, Brazil

February 2022

Eduardo Gil S. Cardoso

Analyzing the Impact of Dimensionality Reduction Over Human Intestinal Absorption Prediction Through Machine Learning Algorithms

Course completion work presented to the Faculty of Computing as one of the requirements to complete the Computer Science bachelor's degree program offered by Universidade Federal do Pará.

Universidade Federal do Pará

Instituto de Ciências Exatas e Naturais

Faculdade de Computação

Supervisor: Prof. Dr. Claudomiro de Souza de Sales Junior

Co-supervisor: Dr. Caio Marcos Flexa Rodrigues

Belém, PA, Brazil

February 2022

Eduardo Gil S. Cardoso

Analyzing the Impact of Dimensionality Reduction Over Human Intestinal Absorption Prediction Through Machine Learning Algorithms/ Eduardo Gil S. Cardoso. – Belém, PA, Brazil, February 2022-

173p. : il. (some color.) ; 30 cm.

Supervisor: Prof. Dr. Claudomiro de Souza de Sales Junior

Course Completion Work – Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Faculdade de Computação, February 2022.

1. dimensionality reduction. 2. machine learning. 2. intestinal absorption. I. Prof. Dr. Claudomiro de Souza de Sales Junior. II. Universidade Federal do Pará. III. Faculdade de Computação. IV. Analyzing the Impact of Dimensionality Reduction Over Human Intestinal Absorption Prediction Through Machine Learning Algorithms

Eduardo Gil S. Cardoso

Analyzing the Impact of Dimensionality Reduction Over Human Intestinal Absorption Prediction Through Machine Learning Algorithms

Course completion work presented to the Faculty of Computing as one of the requirements to complete the Computer Science bachelor's degree program offered by Universidade Federal do Pará.

This work was presented and approved in February 22th, 2022 by a review board encompassing the following members:

Prof. Dr. Claudomiro de Souza de Sales Junior Internal Supervisor — UFPA	**Dr. Caio Marcos Flexa Rodrigues** External Co-supervisor — UFPA
Ing. Msc. Ewerton Cristhian Lima de Oliveira External member — UFPA	**Prof. Dr. Reginaldo Cordeiro dos Santos Filho** Internal member — UFPA

Belém, PA, Brazil
February 2022

Este trabalho é dedicado às pessoas que acreditam, na sua capacidade e de seus próximos.

“Audaces fortuna juvat”

Abstract

A desirable property in drug development is oral delivery. Virtual screening of chemical compounds according to their oral bioavailability with computational intelligence could accelerate the prediction of their human intestinal absorption (HIA). Despite the existence of several studies aimed at predicting the intestinal permeability of chemical compounds, none attempted to evaluate the impact of using physicochemical and structural properties related to oral bioavailability with both dimensionality reduction (DR) and machine learning (ML) techniques. This case study presents an analysis on the impact of applying DR techniques such as Principal Component Analysis (PCA), Kernel PCA (KPCA), Ivis, Truncated Singular Vector Decomposition (TSVD) and Uniform Manifold Approximation and Projection (UMAP) along with ML predictors such as K-Nearest Neighbors (KNN), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM) and Random Forest (RF) in predicting HIA of small molecules, shedding light in the models behavior as dimensionality changes. Results demonstrate that, despite reducing the dimensionality by more than 90%, lower-dimensional models for KNN, SVM and RF still delivered competitive results, demonstrating the viability and potential of projection-based DR as a pre-processing step.

Keywords: projection-based dimensionality reduction, deep learning, drug discovery, human intestinal absorption, machine learning.

Resumo

Uma propriedade desejável no desenvolvimento de drogas é entrega oral. A triagem virtual de compostos químicos de acordo com sua biodisponibilidade oral com inteligência computacional pode acelerar a predição de sua absorção intestinal humana (HIA). A despeito da existência de vários estudos almejando predizer a permeabilidade intestinal de compostos químicos, nenhum tentou avaliar o impacto do uso de propriedades fisico-químicas e estruturais relacionadas à biodisponibilidade oral com técnicas de redução de dimensionalidade (DR) e aprendizado de máquina (ML). Este estudo de caso apresenta uma análise sobre o impacto da aplicação de técnicas de redução de dimensionalidade tais como Análise de Componentes Principais (PCA), PCA baseado em Kernel (KPCA), Ivis, Aproximação e Projeção de *Manifold* Uniforme (UMAP) e Decomposição de Valor Singular Truncado(TSVD), conjuntamente com preditores de ML tais como Redes Neurais Artificiais (ANN), K-Vizinhos mais Próximos (KNN), Máquina de Vetores de Suporte (SVM) e Floresta Aleatória (RF) na predição de HIA de pequenas moléculas, dando foco ao comportamento dos modelos conforme a dimensionalidade varia. Os resultados demonstram que, apesar de reduzir a dimensionalidade em mais de 90%, os modelos de menor dimensionalidade para o KNN, RF e SVM ainda apresentaram resultados competitivos, demonstrando a viabilidade e o potencial de técnicas de DR enquanto etapa de pré-processamento.

Palavras-chave: redução de dimensionalidade baseada em projeção, aprendizado profundo, descoberta de drogas, absorção intestinal humana, aprendizado de máquina.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ADME	absorption, distribution, metabolism, and excretion
ADMET	absorption, distribution, metabolism, excretion, and toxicity
AI	Artificial Intelligence
ANN	Artificial Neural Network
DR	dimensionality reduction
Fcsp^3	Fraction of sp^3-hybridized Carbon Atoms
GI	gastrointestinal
HBA	hydrogen-bond acceptors
HBD	hydrogen-bond donors
HIA	human intestinal absorption
HPO	Hyper Parameter Optimization
HSVM	Hierarchical SVM
KDD	Knowledge Discovery in Databases
KNN	K-Nearest Neighbors
KPCA	Kernel PCA
LDA	Linear Discriminant Analysis
LogP	octanol–water partition coefficient
ML	machine learning
MLP	Multi-Layer Perceptron
MW	Molecular Weight
NAR	number of aromatic rings
NRB	number of rotatable bonds

PCA Principal Component Analysis

PD pharmacodynamic

PK pharmacokinetic

PLS Partial Least Squares

PNN Probabilistic Neural Network

QSAR quantitative structure-activity relationship

RF Random Forest

SVD Singular Value Decomposition

SVM Support Vector Machine

tPSA topological polar surface area

t-SNE t-Distributed Stochastic Neighbor Embedding

TSVD Truncated Singular Vector Decomposition

UMAP Uniform Manifold Approximation and Projection

CONTENTS

CHAPTER

1

INTRODUCTION

1.1 Context and Motivation

Metabolism departments have carved out a niche in modern drug discovery based on the knowledge that successful drugs generally must have proper absorption, distribution, metabolism, and excretion properties (EKINS et al., 2000). Orally-administered drugs, which constitute the preferred route for drug delivery, are subject to all four processes, especially absorption (FEDI et al., 2021). Therefore, gauging the permeability of such compounds figures as a critical task in oral drug formulation (TOZER; ROWLAN, 2016).

One of the significant stages in devising orally-administered drugs is screening for chemical compounds that can cross the intestinal epithelial barrier (SHIN et al., 2018). This concept, which is known as human intestinal absorption (HIA), defines how a given drug candidate crosses the barrier separating the digestive system from the circulatory one, something that decisively influencing the bioavailability of the component in the body (FARRÉ et al., 2020; YOUHANNA; LAUSCHKE, 2021). Given how pivotal drug absorption is in drug discovery, testing the permeability of chemical compounds through organic membranes has become essential in early stages of drug development to achieve the "fail fast, fail early" paradigm (TOZER; ROWLAN, 2016; TA et al., 2021).

A study by Orme (1984) has shown that drugs tend to be absorbed more rapidly from the small intestine than from the stomach, which mostly serves to storage and process incoming food (VERTZONI et al., 2019). For maximized drug bioavailability, solubility in gastrointestinal (GI) fluids and permeability through the epithelial barrier of the small intestine is required (VERTZONI et al., 2019). Both must be proven throughout drug development especially for new chemical entities, whose physicochemical properties are often employed as an indicator of their drug-likeness. As tools to screen for promising compounds, *in vitro* techniques have been developed (JOUBERT et al., 2017).

This notwithstanding, the increase in the number of chemical compounds is such that *in vitro* methods can no longer deal with the high demand of the pharmaceutical industry (DANISHUDDIN et al., 2021). Furthermore, the experimental measurement of pharmacokinetic (PK) properties is still expensive, laborious, and time-consuming (SINGH; GUPTA; BASANT, 2015; BANNIGAN et al., 2021). Consequently, *in silico* testing posed as an alternative to effectively screen for potential drugs and bioactive compounds.

In the realm of pharmaceutics, ML has emerged as a computational approach capable of accelerating and automating analyses of large amounts of data (CARRACEDO-REBOREDO et al., 2021). These and other types of *in silico* testing can be regarded as parts of the computational pharmaceutics field, which integrates Artificial Intelligence (AI) and multi-scale modeling to take advantage of computational resources and algorithms (WANG; YE, et al., 2021). ML has enabled advancements in a variety of fields, including automation of traffic management (NALLAPERUMA et al., 2019), malware detection and classification (GUPTA; RANI, 2020), and drug delivery (HE; LEANSE; FENG, 2021).

Clustering, regression, and classification tend to be recalled first as ML-related tasks; however, dimensionality reduction (DR) also falls under this umbrella. Also broadly employed in statistics, DR can be defined as a means of mapping high-dimensional data in low-dimensional spaces (ESPADOTO et al., 2019; FLEXA et al., 2021). DR compresses the feature space, ultimately producing a representation that better represents data with less redundancy and attenuated multicollinearity that can improve the accuracy of ML algorithms (HOUARI et al., 2016; FLEXA et al., 2021). Such an idea can be further expressed within the concept of intrinsic dimensionality, that when reached, helps achieve a better representation of the data in the building of prediction models (KAK, 2021).

Therefore, this work evaluates the impact of using dimensionality reduction as a pre-processing step in predicting intestinal absorption of small molecules. For such evaluation, five DR algorithms are used, namely Principal Component Analysis (PCA), Kernel PCA (KPCA), Uniform Manifold Approximation and Projection (UMAP), Ivis and Truncated Singular Vector Decomposition (TSVD). These techniques are employed in conjunction with four ML predictors, namely Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN) and Multi-Layer Perceptron (MLP).

1.2 Related Work

This section discusses work on predicting the intestinal absorption of small molecules using ML algorithms, as well as DR applications for assistance in such contexts, whether for visual inspection or feature selection.

Kumar et al. (2017) reduced the dimensionality from 1,529 to 10 molecular descriptors, through means of feature selection such as genetic algorithm and sequential forward

feature selection (SHARMA; KUMAR; VARADWAJ, 2011). The chosen features were collected for 1,242 compounds and given as input for Artificial Neural Network (ANN), KNN, Linear Discriminant Analysis (LDA), Partial Least Squares (PLS), and Probabilistic Neural Network (PNN) classifiers to predict whether a given chemical compound is well-absorbed by the human intestine, with results reaching up to 91% of accuracy.

Esaki et al. (2019) created a three-class HIA prediction model with RF, Linear SVM, and Radial SVM. Feature selection was applied via the `Boruta` package in the `R` programming language to select some of the 7,908 collected descriptors, resulting in feature sets ranging from less than thirty to more than five hundred descriptors depending on the employed ML algorithm. PCA was used to visually analyze how the distribution of the 946 compounds compared against approved drug data.

Lee et al. (2020) chose six descriptors to describe sixty-six molecules. Subsequently, a Hierarchical SVM (HSVM) model was trained to generate a nonlinear quantitative structure-activity relationship (QSAR) model associating these descriptors with intestinal permeability. HSVM was also used by Ta et al. (2021) after applying feature selection via recursive feature elimination in a process that culminated with over a hundred descriptors.

To predict observed influx and efflux permeability (i.e., permeability from the apical side to the basal one and vice-versa, respectively) of chemicals through the intestinal barrier, Kamiya et al. (2021) selected 196 descriptors to describe 219 chemicals. Univariate, bivariate, and trivariate linear regressions, as well as light gradient boosting, were used.

In reviewing some ML applications in precision medicine and drug discovery, Nayarisseri et al. (2021) elaborated on the application of KNN, RF, and SVM classifiers to model absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties such as HIA (YANG et al., 2018), further attesting the pertinence of employing ML classifiers to predict HIA in drug discovery.

In the field of analyzing the impact of DR in the performance of prediction models, Wan, Xu, and Šavija (2021), resorted to PCA to analyse how increasing or reducing the number of principal components affected the cumulative explained variability. This resulted in reducing the dimensionality of the studied data set from eight to six dimensions. No study, however, was performed in terms of model measures such as accuracy to help decide which dimensionality would better optimize predictive performance.

More recently, an article by Stuckman, Walden, and Scandariato (2016) explored this idea in the context of predicting the compressive strength of concrete. In this paper, PCA was used for feature selection, so its projections were matched against prediction models with the whole data set and a lower-dimensional one with manually selected features. The results show the PCA models presenting better performance than the original data set for most cases, but worst when related to the model with manually select data.

1.3 Justification

The performed literary review did not find manuscripts that applied DR via feature extraction prior to the training of models for HIA prediction. Furthermore, no article analyzing the influence of the dimensionality of the projected space on the performance of ML-based models built on them.

To explore both aspects, this work makes use of small molecule data, producing a fifty-dimensional set that is incrementally reduced down to two dimensions. In doing so, this work provides insight on the performance of the applied DR techniques in extracting chemical features as the feature space is compressed.

By resorting to a Hyper Parameter Optimization (HPO) procedure, this work stresses the prediction pipeline from projection to classification in order to get the best result for any given dimensionality-projector-classifier combination, granting that each pipeline has a higher level of competitiveness and fair play. This results in a balanced display of how such models behave as the conditions change.

Ultimately, this work carries implications for drug research and development: DR via feature extraction is seldom applied in this field, and by extensively scrutinizing the implications of doing so, practitioners of the field could get acquainted with and possibly resort to this type of DR as another tool to aid in drug formulation processes. Also, the finding of some dimensionality or pipeline setup with a lighter data set and less redundant information would improve the efficiency in selecting promising compounds.

1.4 Objectives

This work aims at analysing the impact of applying DR techniques in the development of prediction models for intestinal molecular absorption. Within this overreaching goal, there are some specific ones:

- Devise a common pipeline structure that receives the data set as input and returns predictions for whether or not the given sample is absorbed by the intestine, such that only the projectors (if applicable) and classifiers should change between pipelines;
- Choose a representative set of projectors, in this case ivis, KPCA, PCA, TSVD, and UMAP, and classifiers—for which were chosen KNN, MLP, RF, and SVM—that covers traditional and novel techniques;
- Tune the hyper-parameters of each pipeline through Bayesian optimization with cross-validation aiming to maximize the mean accuracy of the pipelines obtained from a cross-validation procedure;

- Perform independent testing of the pipelines after applying the best settings found in the hyperparameter tuning process;
- Analyse the results of the cross-validation and independent testing phases by inspecting the results across the 48 dimensions reached along the process and looking at how they fare against the 50-dimensional baseline for each classifier.

1.5 Structure

The remaining chapters of this work are organized as follows: Chapter 2 elaborates both on the problem domain and on the selection of projectors and classifiers to be studied herein; Chapter 3 clarifies particularities of the data set to be used, the devised pipeline structure, and the process to fine-tune, test, and compare said pipelines from multiple combinations of DR and ML algorithms; Chapter 4 exposes and discusses the obtained cross-validation and independent testing results; and Chapter 5 concludes this manuscript by summarizing its discoveries and enumerating potential future work.

CHAPTER

2

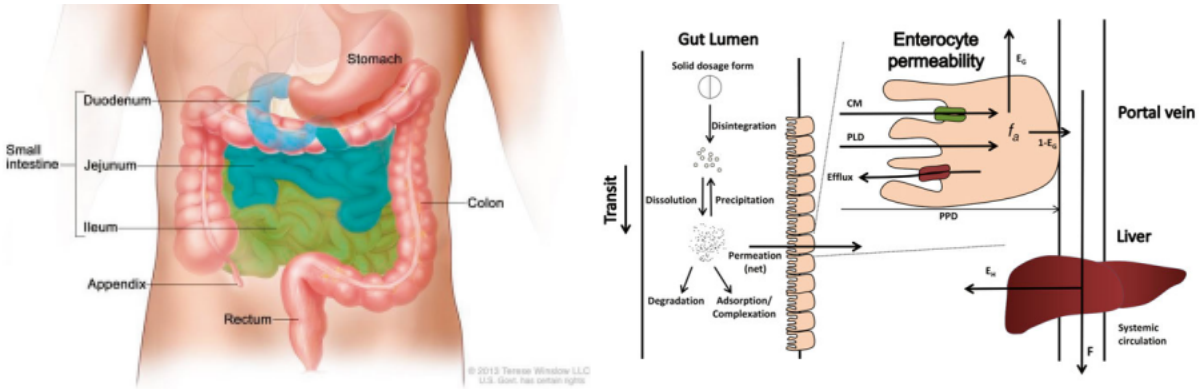
THEORETICAL BACKGROUND

2.1 Drug Absorption

Oral drug treatment is the most common administration route for acting drugs because of tradition, cost-efficient manufacturing, and non-invasiveness (YOUHANNA; LAUSCHKE, 2021; GAVINS et al., 2022). This indicates that the oral route of drug administration will also prevail in the future. However, the GI tract tends to prevent the absorption and translocation of potentially harmful luminal constituents into the central circulation while still allowing the absorption of nutrients and water (TOZER; ROWLAN, 2016). In this context, oral dry absorption is affected by both drug properties and the physiology of the gastrointestinal tract, or patient properties, including drug dissolution from the dosage form, the manner in which drug interacts with the aqueous environment and membrane, permeation across membrane, and irreversible removal by first-pass organs such as the intestine, liver, and lung (MARTINEZ; AMIDON, 2002). Hence, there are several obstacles associated with the oral administration route to overcome for successful systemic drug treatment.

Pharmacokinetics (PKs) can be defined as the study of how an organism affects a drug, and contrasts with pharmacodynamics (PDs), which conversely studies how a drug affects an organism. PKs are determined by four fundamental processes that determine how well a drug will achieve its therapeutic target in the body: absorption, distribution, metabolism, and excretion (ADME) (STORPIRTIS; AL., 2011; FEDI et al., 2021). It is crucial to estimate the intestinal absorption and bioavailability of a compound to understand its PK properties.

To understand the role of ADME processes in orally administered drugs—especially absorption, which is pivotal to maximize bioavailability—, it is relevant to expose the trajectory ensues from ingestion until bioavailability is achieved. The process begins with



(a) Intestine anatomy of a human. (b) Permeability dynamics of HIA.

Figure 1 – Schematic figures describing the location and process of pharmacokinetics and intestinal absorption. Adapted from (DAHLGREN, 2018).

ingestion, which is often coupled with mastication for food. The ingested contents pass through the esophagus, which bridges the mouth with the stomach. Both the mouth and esophagus have surface layers that are not conductive to absorption, meaning that ingested drugs must continue. Figure 1a portrays the involved anatomy, including the stomach, an organ that serves for storage and pre-processing purposes (VERTZONI et al., 2019).

So far, all mentioned parts do not play a role in absorbing elements, but rather in processing it into a more soluble form. Still, drugs must possess relevant properties, such as aqueous solubility, lipophilicity, and hydrophilicity to successfully fulfill these pre-processing steps (FEDI et al., 2021). The stomach, in particular, pushes the food by means of contractions against a closed sphincter termed pylorus until the particles are small enough to pass through it and reach the duodenum, where an absorptive epithelium is encountered. This part, alongside the jejunum and ileum, compose the small intestine, which is a major absorption site for nutrients, liquid, and medicinal products (FEDI et al., 2021; VERTZONI et al., 2019).

Figure 1b shows the process of drug disintegration (e.g., capsule, tablet) in smaller particles, which can be dissolved in GI fluid. The solubility of these compounds in the intestinal fluids must be high enough to provide a fast dissolution rate, and these drug particles must be permeable to cross the apical GI membrane. The drug molecules are considered absorbed after they are transported across the outer lipophilic cell membrane, but this membrane can be resistant to hydrophilic molecules or drugs with high molecular weights (DAHLGREN, 2018). Finally, before its introduction into the central blood circulation, an absorbed drug molecule passes through the liver, where it may be metabolized and lose its pharmacological effect, or be excreted back into the intestines with bile.

It is worth noticing that circulation in the intestine is unique in that the intestine is the anterior or portal tissue that regulates the flow of substrates to the liver. The intestinal venous blood constitutes the majority of the blood supply to the liver, accounting for 75%

of total liver blood flow. For drugs that are highly cleared by the intestine, the contribution of the liver or lung to drug metabolism will become reduced, whereas for drugs that are poorly extracted by the intestine, the substrate is able to reach the next organs, the liver and the lung, for removal (HIRAYAMA; XU; PANG, 1989; GUGLER et al., 1975).

2.2 Machine Learning and Dimensionality Reduction

At an abstract level, the Knowledge Discovery in Databases (KDD) field is concerned with the development of methods and techniques for making sense of data. Historically, the notion of finding useful patterns in data has been given a variety of names, including data mining, knowledge extraction, information discovery, information harvesting, data archaeology, and data pattern processing (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). In this context we have machine learning, which can be defined as a branch of AI that aims at developing and applying computer algorithms that learn from raw data, in order to later perform a specific task (CARRACEDO-REBOREDO et al., 2021), and that is where we come accross Dimensionality Reduction.

DR can be achieved by selecting or extracting features. Selecting features can be done by filtering features agnostically to the ML algorithm used (LI et al., 2012; WAN; WANG, et al., 2016) or by wrapping—i.e., appending variables to others or removing them according to prediction measures (e.g., accuracy) (BRUNELLO et al., 2019; PEREIRA et al., 2018). However simple, there are some drawbacks to this approach: selecting a subset of a large feature set might involve testing an exponential number of possible feature combinations. Moreover, this method might cause loss or distortion of data structures and arrangements (FLEXA et al., 2021).

As an alternative, DR can be achieved by extracting features of data. Assuming $\mathcal{X} \in \mathbb{R}^D$ to be a D-dimensional data set describing N objects, a DR technique that extracts features learns a mapping function $\mathcal{P}(\mathcal{X}) \rightarrow \mathcal{Y}$, where $\mathcal{Y} \in \mathbb{R}^d$ is a representation of $\mathcal{X}$ in d dimensions such that $d < D$. This kind of DR algorithm, which is the one used in this work, is called projector ($\mathcal{P}$). Its resulting representation $\mathcal{Y}$ is termed projection (ESPADOTO et al., 2019).

There are three relevant aspects in terms of which projectors can be classified: neighborhood, transformation, and supervision. $\mathcal{P}$ can preserve local or global neighborhoods. Local techniques preserve properties related to an instance considering its nearest neighbors. They tend to promote greater inter-cluster separation in detriment of global data arrangements. Conversely, global techniques promote fidelity by bringing properties observed in the original dimensionality to the produced projections (ESPADOTO et al., 2019). $\mathcal{P}$ can also apply linear or non-linear transformations. Linear techniques tend to be easier to understand and implement, but cannot depict complex structures. Contrastingly,

non-linear algorithms are more powerful, but customarily entail costlier routines and the tuning of input hyper-parameters (ESPADOTO et al., 2019). Lastly, $\mathcal{P}$ can be supervised or unsupervised. Supervision refers to the provision of labels to the algorithm alongside the data. Supervised techniques might take advantage of the labels to better fit the data; however, it is not always feasible to attain data labels. Contrariwise, unsupervised techniques only rely on data-intrinsic information properties.

2.2.1 Projectors

2.2.1.1 Principal Component Analysis

Arguably the most popular DR technique, PCA is a global, linear, unsupervised projector that is hyper-parameter free and works by converting a data set of possibly correlated variables with orthogonal transformations into a set of linearly uncorrelated principal components (JOLLIFFE, 2002; ESPADOTO et al., 2019). In doing so, this algorithm attempts to maximize within-data variation (FLEXA et al., 2021). Variations of this algorithm include Kernel PCA and Incremental PCA (ESPADOTO et al., 2019).

2.2.1.2 Truncated Singular Value Decomposition

TSVD acts as a low-rank matrix approximation, much like PCA, and is, as the name implies, a truncated version of the Singular Value Decomposition (SVD) (HALKO; MARTINSSON; TROPP, 2009). Overall, TSVD is considered a method for regularization of ill-posed linear least squares problems, and is used to define new a well-posed problem, with a solution which is less sensitive to perturbations (HANSON, 1971).

2.2.1.3 Kernel Principal Component Analysis

As mentioned in the previous subsection, KPCA is a variation of the PCA approach, that applies Kernel-based methods like the ones found on SVMs (SCHÖLKOPF; SMOLA; MÜLLER, 1998). Instead of calculating the covariance of matrix, KPCA computes the principal eigen vectors of the kernel matrix, which enables it to extract nonlinear principal components using less computation power. (AYESHA; HANIF; TALIB, 2020)

2.2.1.4 Uniform Manifold Approximation and Projection

UMAP is an iterative, neighborhood-based projector that relies on a theoretical framework based in Riemannian geometry and algebraic topology (MCINNES; HEALY; MELVILLE, 2018). Unlike the random initialization employed by default in t-Distributed Stochastic Neighbor Embedding (t-SNE), UMAP initializes the projection with a technique termed Spectral Embedding. Moreover, it employs cross-entropy as its objective function

instead of Kullback-Leibler divergence like t-SNE. These changes amount to faster run times and better learning of global and local arrangements compared to t-SNE.

2.2.1.5 Ivis

Introduced under the context of single-cell analysis, Ivis is a neighborhood-based non-linearDR technique that reduces the dimensionality of data by means of a Siamese neural network (SNN) (KOCH; ZEMEL; SALAKHUTDINOV, et al., 2015) with a triplet loss function. It aims to capture local and global features of high-dimensional data sets in a scaling manner, and can produce supervised and unsupervised projections (SZUBERT et al., 2019).

2.2.2 Classifiers

2.2.2.1 K-Nearest Neighbors

K-Nearest Neighbors (KNN) was first defined by Fix and Hodges (1951) as a means of tackling the discrimination problem when no information on class distributions are available. KNN is a popular non-supervised algorithm that can be used for classification and regression tasks. Its underlying assumption is that the neighborhood of a given instance, for which a label is wanted, can deliver useful information (KRAMER, 2013).

2.2.2.2 Multi-Layer Perceptron

MLP is an example of an ANN that is used extensively for the solution of a number of different problems, including pattern recognition and interpolation (NORIEGA, 2005). As the name suggests, MLP consists of multiple layers of simple, two-state, sigmoid processing elements, the so called perceptrons, that interact with each other using weighted connections (PAL; MITRA, 1992).

2.2.2.3 Support Vector Machine

Support Vector Machine (SVM) is a ML algorithm based on statistical-learning theory (JAMES et al., 2013). It focuses on boundaries between classes and maps the input space created by independent variables using a nonlinear transformation according to a kernel function.

2.2.2.4 Random Forest

Random Forest (RF) is an ensemble technique that utilizes several independent decision trees to perform classification and regression (JAMES et al., 2013). Additionally, RF is also appealing for its low hyperparameter count, outlier detection and ease in dealing with higher-dimensional data (CUTLER; CUTLER; STEVENS, 2012).

2.2.3 Bayesian Optimization

Bayesian optimization is often referred to as a good option for hyperparameter tuning in unknown circumstances in continuous domains (FRAZIER, 2018), as the method starts assuming a simple Gaussian distribution for the surrogate model and then improves its model as the results are gathered and analysed in means of performance, thus figuring as a contained self incrementing method (SNOEK; LAROCHELLE; ADAMS, 2012), that lacks the need for discrete search steps, instead covering a space simply defined by upper and lower bounds.

CHAPTER

3

METHODOLOGY

The brute data used for this work consists of a series of `SMILES` descriptors, one for each small molecule to be analyzed. Physicochemical properties and organic structures are extracted from these descriptors and passed to a DR technique to reduce their dimensionality. Lastly, the resulting data set is used to train a ML model to return binary predictions on the HIA of a small molecule.

Prior to the independent testing, key classifier hyper-parameters are adjusted using Bayesian search with 10-fold cross-validation on the training set. Then, twenty executions are made with different train-test-splits and the same classifier hyper-parameters, from which key model measures are recorded and averaged.

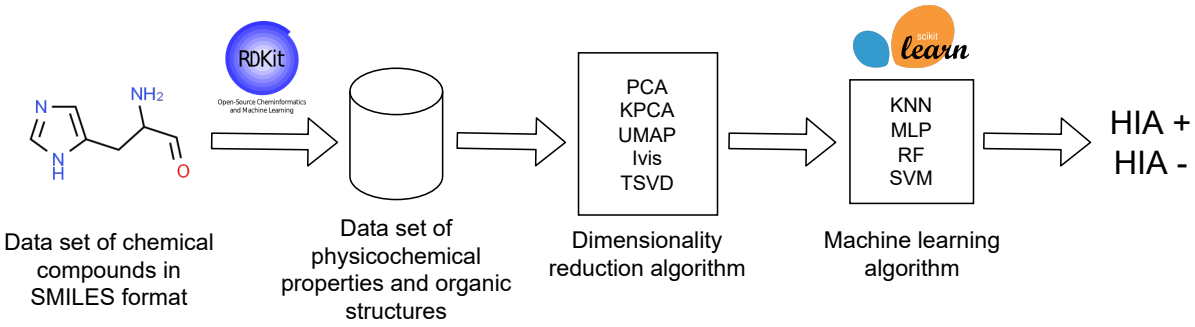


Figure 2 – Schematic of DR/ML pipeline for HIA prediction.

3.1 Data Set

The data set utilized in this paper for HIA is composed of 304 small molecules tested and evaluated for intestinal permeability in (WANG; AL, 2016). The data set contains 304 compounds with 152 molecules that are absorbed (HIA +) and 152 that are not (HIA -). The feature set used includes eight physicochemical and 42 structural

properties related to chemical compound absorption extracted from molecules in `SMILES` format using `RDKit` (LANDRUM, 2020) in `Python`.

The physicochemical ones are Molecular Weight (MW), octanol–water partition coefficient (LogP), topological polar surface area (tPSA), hydrogen-bond donors (HBD), hydrogen-bond acceptors (HBA), number of rotatable bonds (NRB), Fraction of sp^3-hybridized Carbon Atoms ($Fcsp^3$), and number of aromatic rings (NAR). The structural ones are listed below.

- aliphatic carboxylic acids;
- aliphatic hydroxyl group;
- N functional groups attached to aromatics;
- aromatic carboxylic acids;
- aromatic nitrogens;
- aromatic amines;
- aromatic hydroxyl;
- carboxylic acids;
- carbonyl O;
- thiocarbonyl;
- imines;
- primary amines;
- secondary amines;
- tertiary amines;
- hydroxylamine;
- thiol;
- aldehydes;
- amides;
- amidine;
- anilines;
- azide groups;
- azo groups;
- benzene rings;
- bicyclic;
- esters;
- guanidine;
- halogens;
- ketones;
- nitriles;
- nitro groups;
- phenols;
- urea groups;
- primary amides;
- ether oxygens;
- furan rings;
- thioether;
- sulfonamides;
- sulfone groups;
- quaternary nitrogen;
- methoxy groups;
- beta lactams; and
- cyclic esters (lactones).

3.2 Bayesian Optimization

As mentioned before, this work makes use of the DR techniques PCA, KPCA, Ivis, TSVD and UMAP for projecting the original 50-dimensional data set into increasingly low dimensions, along with the ML algorithms KNN, MLP, RF and SVM as classifiers for providing prediction models in the means of predict the whether a given molecule is absorbed or not by the intestine. In order to do this in a more uniform basis, this work uses Bayesian search for optimizing each projector-classifier pair by tuning its hyperparameters when necessary, in order to provide optimal performance for each given dimensionality to which the original data set is projected to. In this context, Table 1 presents how the search space was set for each projector and classifier used.

Table 1 – Search space of the considered hyper-parameters for all estimators of the projection and classification phases.

Search space		Hyper-parameter	Type	Interval
Phase	Estimator			
projection	PCA	—	—	—
	TSVD	`algorithm`	categorical	[`arpack`]
	KPCA	`eigen_solver`	categorical	[`dense`]
		`gamma`	real	[0, 1]
		`kernel`	categorical	[`cosine`, `linear`, `rbf`, `sigmoid`]
		`n_jobs`	categorical	[`-1`]
	UMAP	`n_epochs`	integer	[100, 2000]
		`n_neighbors`	integer	[100, 2000]
		`random_state`	categorical	[`2021`]
		`target_weight`	real	[0, 1]
	Ivis	`k`	integer	[3, 103]
		`model`	categorical	[`hinton`, `maaten`, `szubert`]
		`n_epochs_without_progress`	integer	[10, 50]
		`supervision_weight`	real	[0, 1]
classification	KNN	`leaf_size`	integer	[3, 30]
		`n_neighbors`	integer	[3, 30]
		`p`	integer	[1, 5]
		`weights`	categorical	[`distance`, `uniform`]
	MLP	`activation`	categorical	[`logistic`, `relu`, `tanh`]
		`alpha`	real	[1e-4, 1e-2]
		`hidden_layer_sizes`	(integer, integer)	([50, 150], [50, 150])
		`max_iter`	integer	[100, 2000]
		`random_state`	categorical	[`2021`]
	SVM	`C`	real	[1e-3, 1e3]
		`gamma`	real	[1e-2, 1e2]
		`kernel`	categorical	[`linear`, `rbf`, `sigmoid`]
		`max_iter`	categorical	[`100000`]
		`probability`	categorical	[`True`]
		`random_state`	categorical	[`2021`]
	RF	`max_depth`	integer	[3, 15]
		`n_estimators`	integer	[50, 150]
		`n_jobs`	categorical	[`-1`]
		`random_state`	categorical	[`2021`]

CHAPTER

4

RESULTS AND DISCUSSION

This chapter explores the proposed methodology's results and is divided in two sections. Section 4.1 displays the results found in the cross-validation phase of the proposed pipeline, while Section 4.2 discuss the results of the independent tests performed on the models achieved by the Bayesian optimization.

As one may suppose, experimenting with twenty possible pipelines for each of the forty-eight evaluated dimensionalities results in a plethora of information. Showing it here in its raw form would be rather extensive, which is why the produced results in their integral form are in the appendices of this manuscript.

4.1 Cross Validation Results

As discussed in Chapter 3, this work relies on Bayesian optimization and 10-fold cross-validation to provide optimal hyperparameter settings for each produced model. The results of this process are displayed herein.

Figure 3 presents the execution time for the fitting of the best models found by the HPO, arranged either by classifier or by projector. From Figure 3.a, we notice straight away the high fitting time for the UMAP projector, which reaches up to five seconds while the other three projectors take within one second. Due to visualization issues, the fitting times for models involving the use of the Ivis projector have been removed from Figure 3 since its computationally-intensive neural network processing significantly impacts its execution time, thus dwarfing the remaining projections even for logarithmic scales. This is compensated by Figure 4, which compares Ivis in terms of average runtime against UMAP, whose runtime is the highest of all projectors other than Ivis.

Figure 5 displays the fitting times for the models as the dimensionality of the projections increases. It is possible to notice a subtle relation between the fitting time and

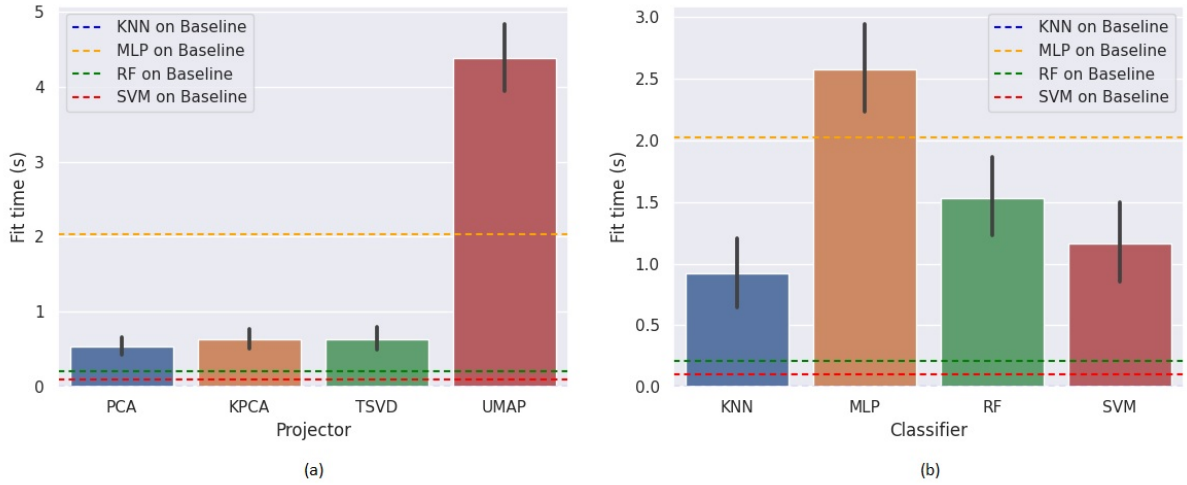


Figure 3 – Fit time for cross validation arranged by classifier and projector.

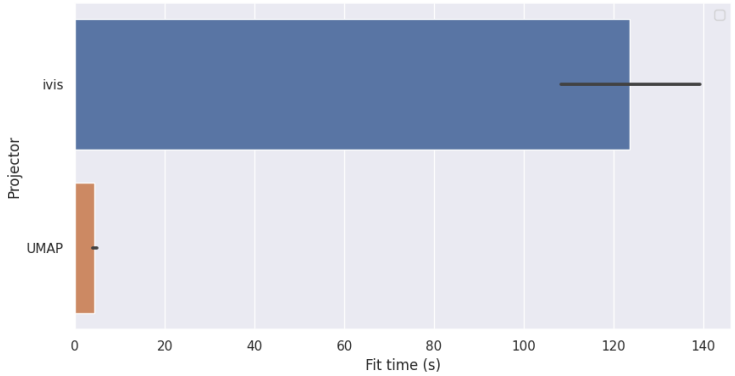


Figure 4 – Average fit times for ivis and UMAP projections

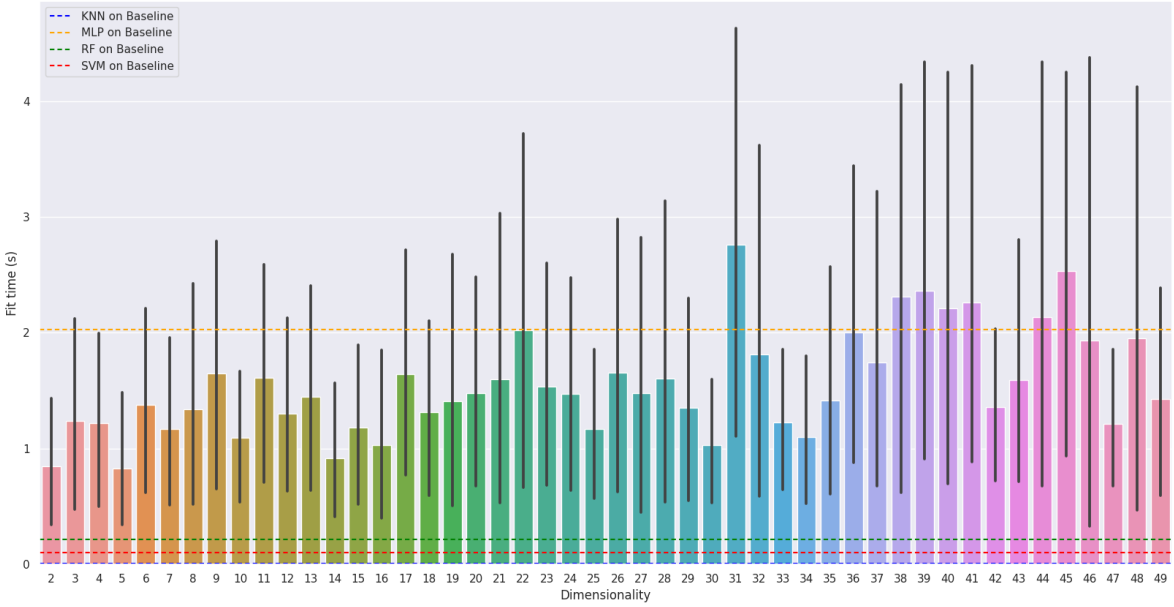


Figure 5 – Fit time for cross validation arranged by dimensionality.

the dimensionality for the times seem to get higher as the dimensionality does, along with some period between peaks.

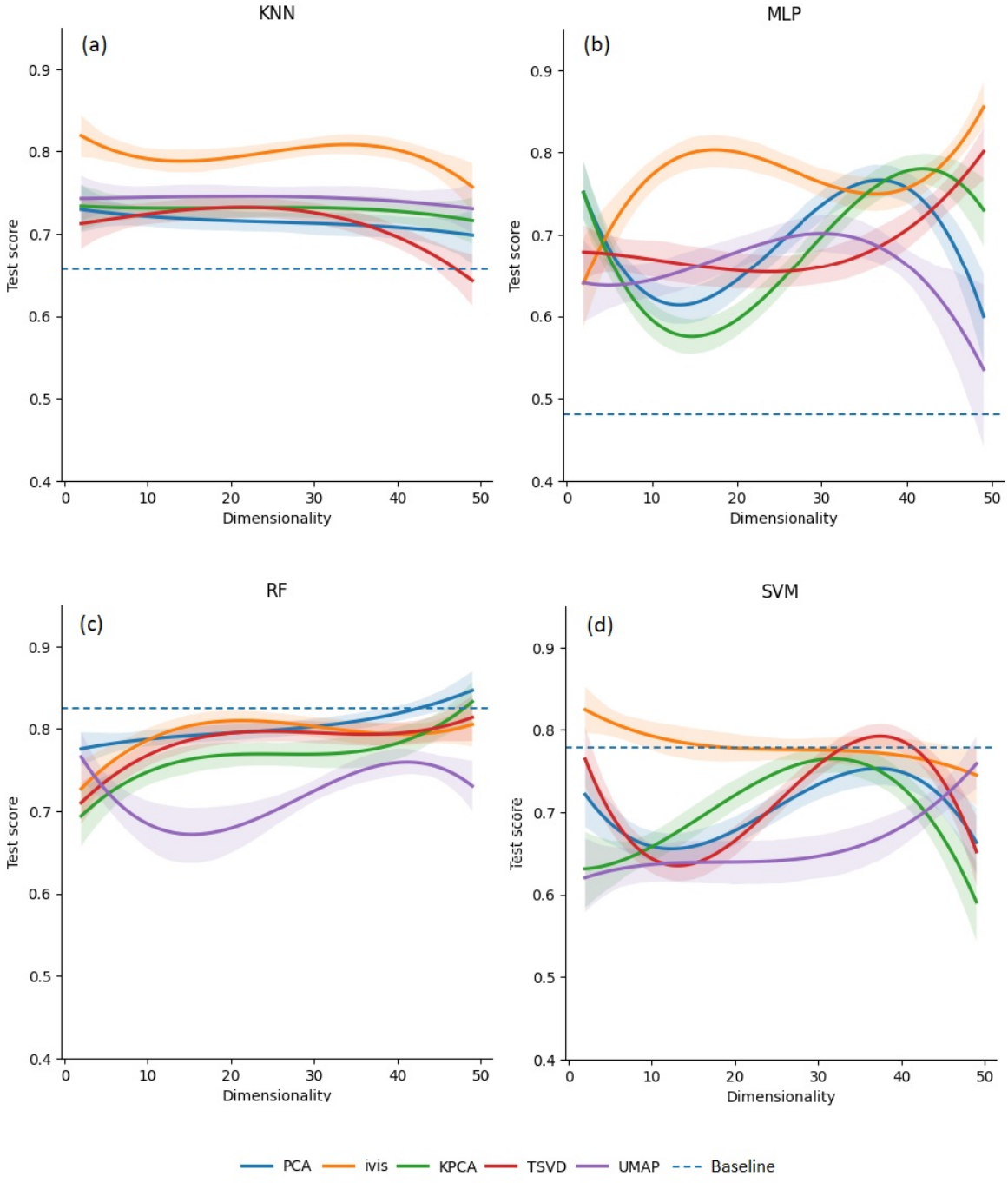


Figure 6 – Test results for cross validation.

Now examining the scores reached by the best models, Figure 6 samples the score distributions as third-degree polynomial regression lines. Each line represents a projector, whereas each plot depicts results for a different classifier. From this figure, we may perceive in Figure 6.a the good performance presented by the KNN models, which show a high level of consistency, with slight variation noticeable for the PCA, KPCA and UMAP projections, along with high overall scores, with only models with TSVD projections going

under baseline scores for higher dimensionalities. Such result is quite appealing when considering the low fitting time for the KNN models.

From Figure 6.b, however, it is noticeable the instability of the MLP-based models, which present a high variation, ranging from a score of 0.5 up to almost 0.9, likely due to unfortunate weight distribution between perceptrons, leading up to a poor performance. Such assumption is supported by Figure 7, which shows the distribution for the max number of iterations utilized by the achieved MLP models during the performed cross-validation.

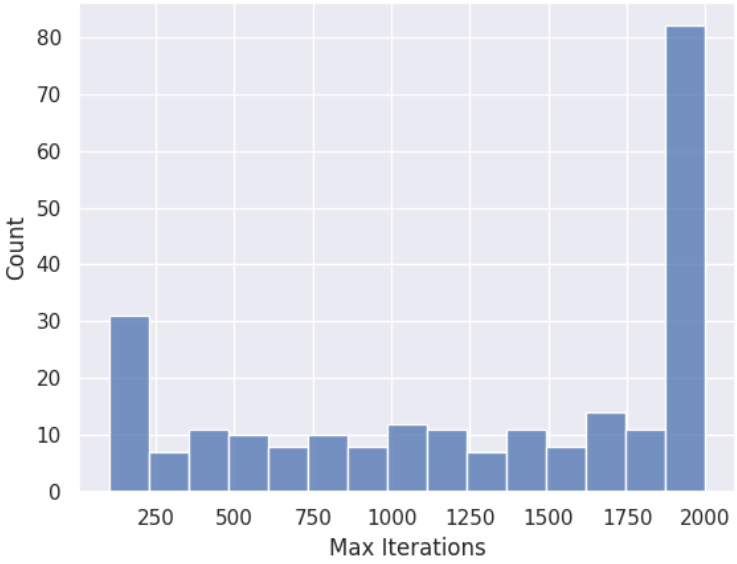


Figure 7 – MLP Max Iterations parameter histogram

The histogram presented in Figure 7 suggests that the selected upper bound for the Bayesian optimization regarding the MLP classifier impeded the exploration of higher numbers of iterations, possibly leading models to premature convergence due to lack of time for further exploration. This is supported by the fact that a high amount of models chose the upper bound of the search space as the optimal setting for this hyperparameter.

Back in Figure 6.c, it is also worth noticing the scores for the RF models, which figure consistently higher than the other classifiers scores except for models with UMAP projections, which show more deviance and slightly worse performance. Moreover, the RF models scores also present a noticeable direct proportionality with the model's projection dimensionality, again except for UMAP-based models.

By inspecting the hyperparameter settings of UMAP for the maximum number of iterations, we find a similar case to what happened to the MLP models, as portrayed by Figure 8. We may once again notice a high amount of models that chose the upper bound of the search space to set this hyperparameter. As with the MLP case, this might imply that the stipulated search space was not optimal for a proper optimization here as well,

thus leading to premature performance and, ultimately, to unstable models that do not quite reach all of UMAP's feature extraction potential.

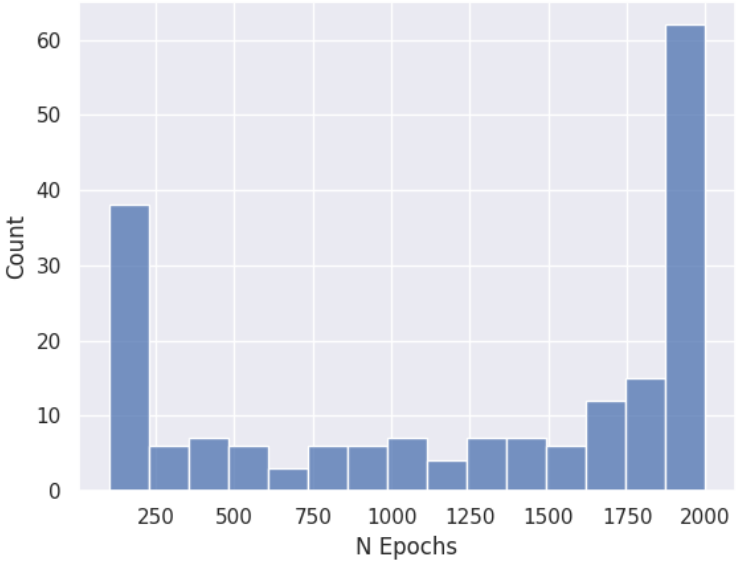


Figure 8 – UMAP N Epochs parameter histogram

4.2 Independent Testing Results

Regarding the independent testing stage, Figure 9 portrays how the average accuracy across all pipelines with DR is affected as the dimensionality increases. No perceivable tendency tying accuracy with dimensionality can be visually drawn. This behavior is likely

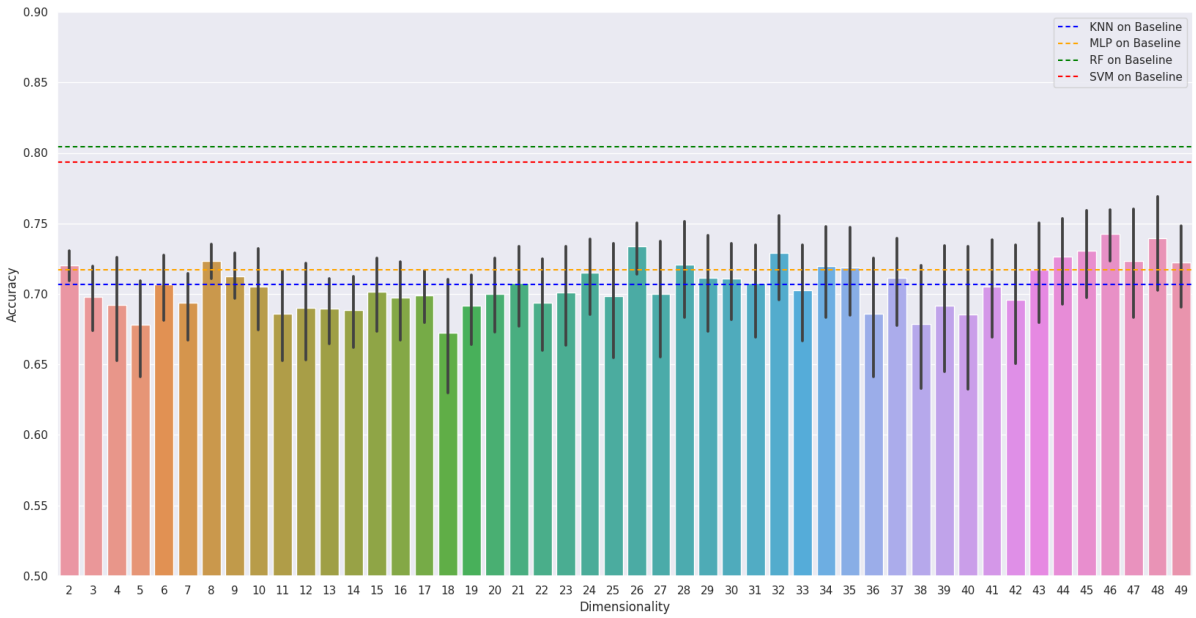


Figure 9 – Average accuracy distribution for independent testing sorted by dimension

due to the Bayesian optimization applied, that is tuning the hyperparameters for each projector granting optimal conditions for each given dimensionality-projector-classifier scenario, as one might suggest by examining the high variation in hyperparameter values presented in Table 4.

Furthermore, when subjected to independent testing, the optimized models achieved performance with somewhat similar behavior to the test scores displayed in the previous section, as noticed in Figure 10. Exceptions are the SVM models in Figure 10.d, which

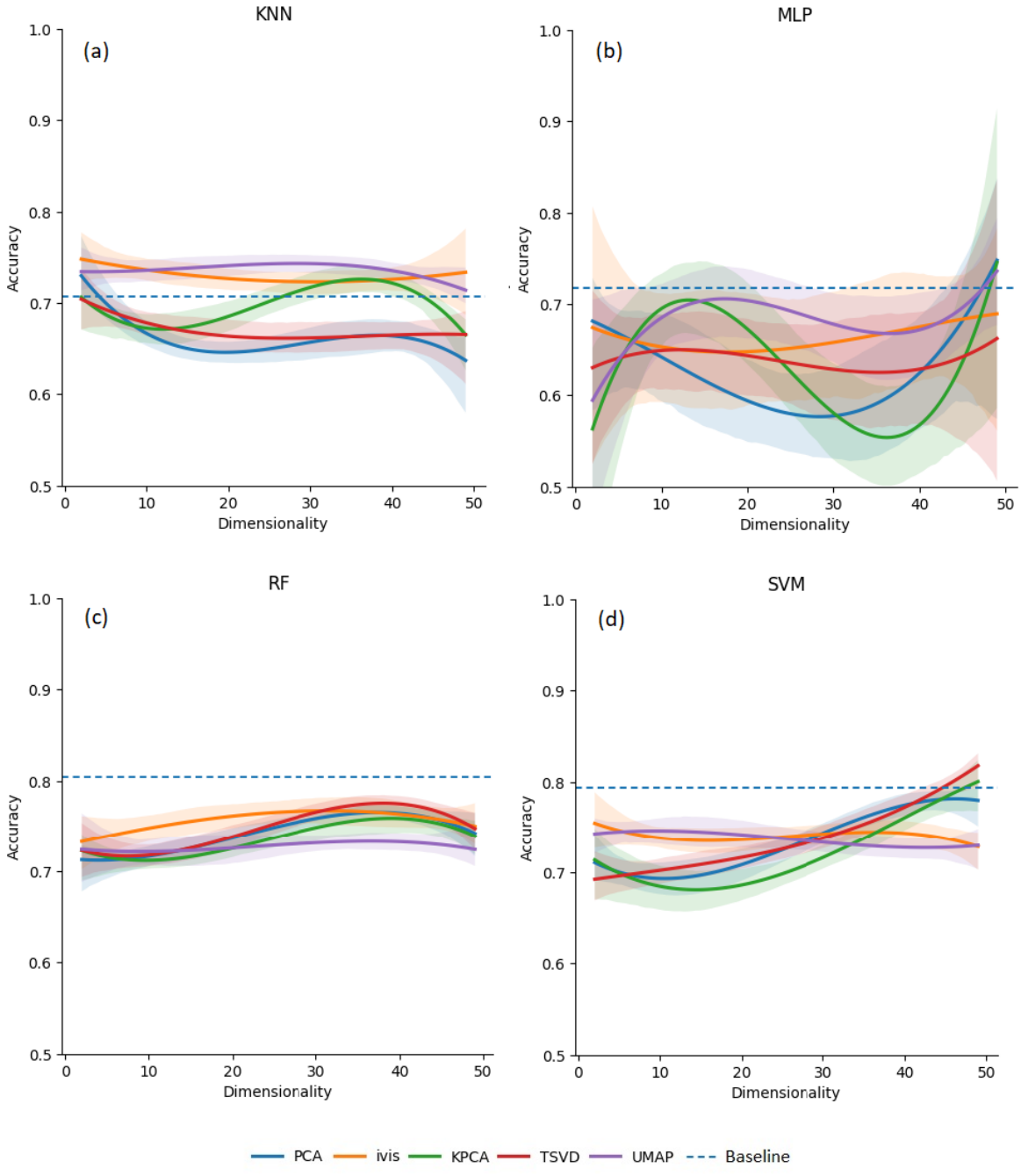


Figure 10 – Accuracy distribution for independent testing

present much higher consistency and even a somewhat strong positive correlation with the dimensionality of its projections. For higher-dimensional spaces, SVM even manages to beat the performance of the baseline model, which uses the original data set.

In Figure 10.a, we have the KNN models with DR outperforming the baseline model with up to 5% higher accuracy. It is also worth noticing in the KNN plot how much better the lower-dimension models performed for most projections.

Looking now at the accuracy of RF and KNN models, we may notice a compelling commonplace for lower-dimensional models, as their accuracy seems to reside consistently in the 70-75% accuracy range, which roughly extends for the SVM models. Such results present a considerably better performance when compared to the baseline accuracy reached by KNN and MLP but fall behind when matched against the baseline accuracy for RF and SVM models.

However, even though the results reached by these lower-dimensional projections were lower than those reached by the given baselines, the fact that the accuracy loss barely reaches 10% shows how useful such models are in reducing the computational load with an acceptable loss in performance. For instance, if we were to apply such prediction models into a data set with more indiscriminate descriptors, regardless of their possible redundancy, we would be able to tone down the computational cost of the prediction algorithms by applying the proposed DR techniques as set by the performed optimization while maintaining acceptable overall accuracy, thus enabling for more elaborate exploration and, for two and three-dimensional models, visual inspection.

Additionally, in Figure 10.b, we may once again perceive the odd behavior of the MLP models, with high instability and low overall accuracy, caused by the circumstances presented in Figure 7, and reflecting in the oscillating specificity and sensitivity achieved

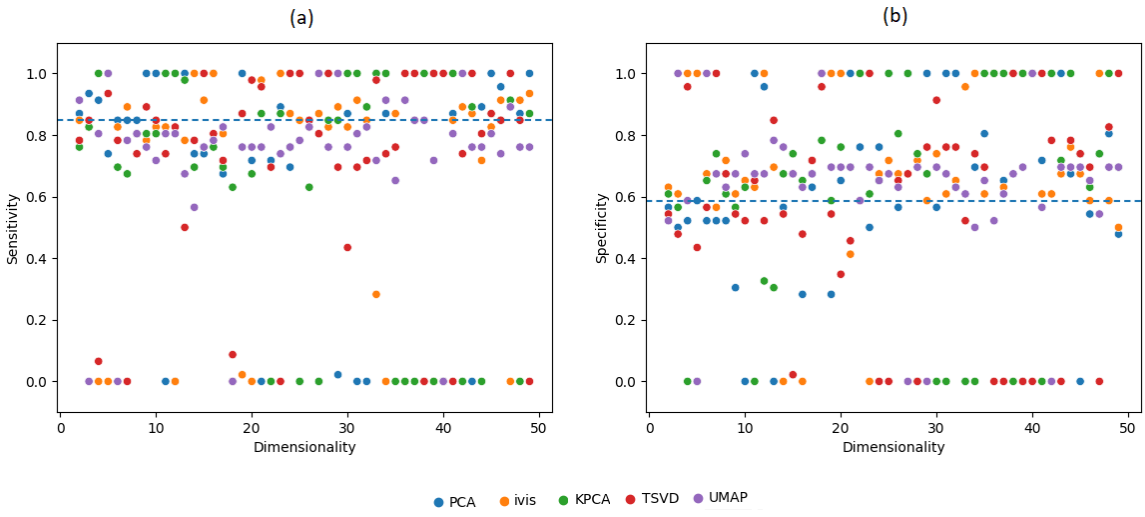


Figure 11 – Sensitivity and Specificity for MLP models

in the independent test step by such models, as shown in Figure 11. We can see the lines formed by the models metrics at the points of 0% and 100%, both for sensitivity and specificity, which implies that such discrete models have come to a conjecture where the threshold for defining whether or not the givens samples have acceptable HIA is either so low that every sample will be accepted, hence, the 100% sensitivity cases; or is so high that no samples are able to be accepted, thus producing the 100% specificity models .

Figure 12 presents the accuracy distribution once again related to dimensionality, but this time sorted by the used classifier in Figure 12.a and by applied projector in Figure 12.b. From these plots, we may notice how tightly related the accuracy for different projections seems to be compared with models with different classifiers, which remain that previous mentioned lower-dimensional commonplace, but start diverging as the dimensionality gets past ten features.

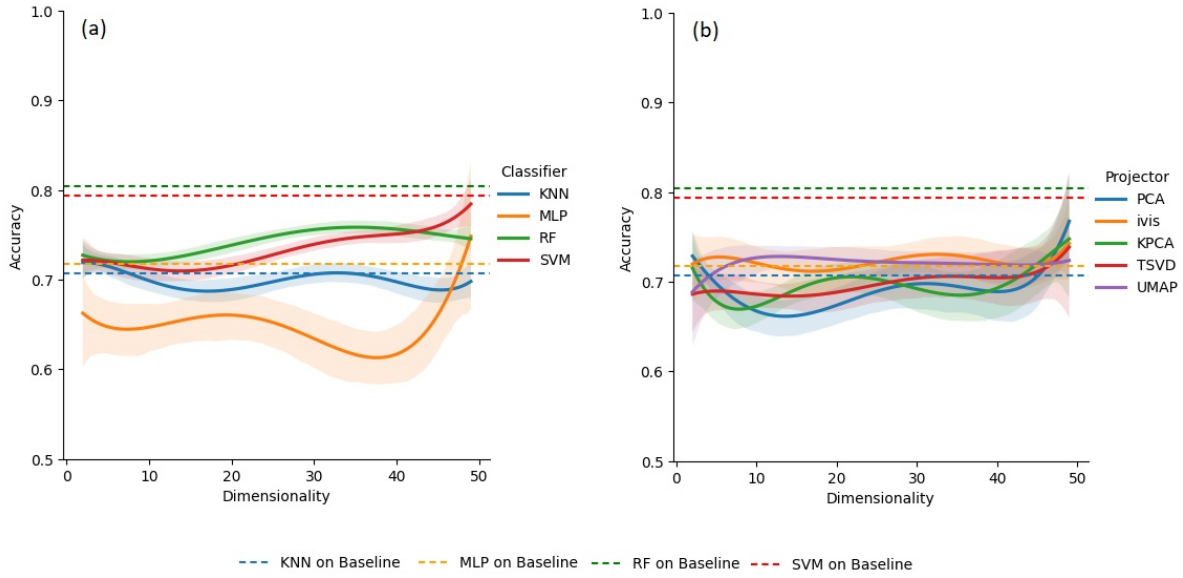


Figure 12 – Classifier and Projector accuracy distribution for independent testing

Hence, we can imply, given the visual inspection of Figure 12, that the classifier choice is much more critical for the model performance than the projector choice, as the latter seems to differ little as the dimensionality varies.

CHAPTER

5

CONCLUSION AND FUTURE STUDIES

Altogether, this work presented results that grant a deep analysis of the overall behavior of HIA prediction models when submitted to DR techniques, approaching different fronts of the developed pipeline, bringing together its particularities for a complete understanding of its results and possible tweaks for further improvement. In this sense, after analysing the results presented in Chapter 4, we can get into a few significant conclusions, as laid in this chapter.

There is a slight correlation between the fitting time of the models and their dimensionality, as they seem to be directly proportional. This implies that it is preferable to choose for projecting the available data sets to lower dimensions, thus reducing the computational time of its fitting process and further improving performance. There seems to be a commonplace for low-dimensional models accuracy, which floats around 70-75% for the given data set. As mentioned in Section 4.2, this result sheds light on the low feature loss provided by the applied DR techniques, which manage to reduce the dimensionality of the data set while maintaining its characteristics and its models' accuracy, granting lower computational load and the possibility of visual inspection.

The classifier choice for the model is considerably more impactful for its performance than the projector choice. As shown in the previous chapter, the accuracy distributions seem to vary relatively little from projector to projector, preserving some global trend, while the accuracy distributions o each classifier grant more distinct behavior, perhaps except for the RF-SVM pair, which share some similarity.

Given the previous item, it is pertinent to give more thought to the classifier choice in real-world scenarios, maybe favoring lower computing times, for which KNN would be a good choice, or preferring better performance, for which case SVM would be a more suitable choice. Subsequently, a projector could be picked from the considered ones with good confidence that the impact of this choice would be lesser than the one of the classifier.

Furthermore, some points were found in the developed pipeline where the search bounds of the Bayesian optimization procedure seem to have constrained the optimization of certain techniques, such as the projector UMAP and the classifier MLP. Hence, it would be appropriate to consider developing follow-up studies extending this said boundaries for allowing such techniques to achieve their full potential. Of course, such work would also be coupled with an analysis on the computational times for such extended optimization, along with an inspection of how the obtained performance metrics stand against the results presented herein.

As a proof of concept, this work handles a relatively small data set, in order to be able to manage its robust pipelines and its optimization, which, even for such small sampling, was already hitting up to one million models searched, each with their own intrinsic extensive computations. Therefore, it would make sense to perform further exploration for higher-dimensional data sets on HIA and, in doing so, make use of more suitable settings for the developed prediction models given the information granted by this work, thus narrowing down the hyperparameter space.

REFERENCES

AYESHA, Shaeela; HANIF, Muhammad Kashif; TALIB, Ramzan. Overview and comparative study of dimensionality reduction techniques for high dimensional data. **Information Fusion**, Elsevier, v. 59, p. 44–58, 2020. Cit. on p. 22.

BANNIGAN, Pauric et al. Machine learning directed drug formulation development. **Advanced Drug Delivery Reviews**, v. 175, p. 113806, Aug. 2021. ISSN 0169409X. DOI: 10.1016/j.addr.2021.05.016. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0169409X21001800>. Visited on: 22 Jan. 2022. Cit. on p. 15.

BRUNELLO, Andrea et al. Multiobjective evolutionary feature selection and fuzzy classification of contact centre data. **Expert Systems**, v. 36, Mar. 2019. Cit. on p. 21.

CARRACEDO-REBOREDO, Paula et al. A review on machine learning approaches and trends in drug discovery. **Computational and Structural Biotechnology Journal**, v. 19, p. 4538–4558, 2021. ISSN 20010370. DOI: 10.1016/j.csbj.2021.08.011. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2001037021003421>. Visited on: 22 Jan. 2022. Cit. on pp. 15, 21.

CUTLER, Adele; CUTLER, D Richard; STEVENS, John R. Random forests. In: ENSEMBLE machine learning. [S.l.]: Springer, 2012. P. 157–175. Cit. on p. 23.

DAHLGREN, David. **Biopharmaceutical aspects of intestinal drug absorption: Regional permeability and absorption-modifying excipients**. 2018. PhD thesis – UPPSALA UNIVERSITY. Cit. on p. 20.

DANISHUDDIN et al. A decade of machine learning-based predictive models for human pharmacokinetics: Advances and challenges. **Drug Discovery Today**, s1359644621004074, Sept. 2021. ISSN 13596446. DOI: 10.1016/j.drudis.2021.09.013. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1359644621004074>. Visited on: 22 Jan. 2022. Cit. on p. 15.

EKINS, Sean et al. Present and future in vitro approaches for drug metabolism. **Journal of pharmacological and toxicological methods**, Elsevier, v. 44, n. 1, p. 313–324, 2000. Cit. on p. 14.

ESAKI, Tsuyoshi et al. Constructing an In Silico Three-Class Predictor of Human Intestinal Absorption With Caco-2 Permeability and Dried-DMSO Solubility. **Journal of Pharmaceutical Sciences**, v. 108, n. 11, p. 3630–3639, Nov. 2019. ISSN 00223549. DOI: 10.1016/j.xphs.2019.07.014. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0022354919304496>. Visited on: 23 Jan. 2022. Cit. on p. 16.

ESPADOTO, Mateus et al. Towards a quantitative survey of dimension reduction techniques. **IEEE transactions on visualization and computer graphics**, IEEE, 2019. Cit. on pp. 15, 21, 22.

FARRÉ, Ricard et al. Intestinal Permeability, Inflammation and the Role of Nutrients. **Nutrients**, v. 12, n. 4, p. 1185, 23 Apr. 2020. ISSN 2072-6643. DOI: 10.3390/nu12041185. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7231157/>. Visited on: 1 Feb. 2022. Cit. on p. 14.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**, ACM New York, NY, USA, v. 39, n. 11, p. 27–34, 1996. Cit. on p. 21.

FEDI, Arianna et al. In vitro models replicating the human intestinal epithelium for absorption and metabolism studies: A systematic review. **Journal of Controlled Release**, v. 335, p. 247–268, July 2021. ISSN 01683659. DOI: 10.1016/j.jconrel.2021.05.028. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0168365921002522>. Visited on: 24 Jan. 2022. Cit. on pp. 14, 19, 20.

FIX, Evelyn; HODGES, JL. **Nonparametric Discrimination: Consistency Properties**. Randolph Field, Texas, Feb. 1951. Available from: <https://apps.dtic.mil/dtic/tr/fulltext/u2/a800276.pdf>. Cit. on p. 23.

FLEXA, Caio et al. Polygonal Coordinate System: Visualizing high-dimensional data using geometric DR, and a deterministic version of t-SNE. **Expert Systems with Applications**, Elsevier, v. 175, p. 114741, 2021. Cit. on pp. 15, 21, 22.

FRAZIER, Peter I. Bayesian optimization. In: RECENT advances in optimization and modeling of contemporary problems. [S.l.]: Informs, 2018. P. 255–278. Cit. on p. 24.

GAVINS, Francesca K.H. et al. Machine learning predicts the effect of food on orally administered medicines. **International Journal of Pharmaceutics**, v. 611, p. 121329, Jan. 2022. ISSN 03785173. DOI: 10.1016/j.ijpharm.2021.121329. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0378517321011352>. Visited on: 22 Jan. 2022. Cit. on p. 19.

GUGLER, ROLAND et al. Pharmacokinetics of drugs in patients with the nephrotic syndrome. **The Journal of clinical investigation**, Am Soc Clin Investig, v. 55, n. 6, p. 1182–1189, 1975. Cit. on p. 21.

GUPTA, Deepak; RANI, Rinkle. Improving malware detection using big data and ensemble learning. **Computers & Electrical Engineering**, v. 86, p. 106729, Sept. 2020. ISSN 00457906. DOI: 10.1016/j.compeleceng.2020.106729. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S004579062030584X>. Visited on: 4 Apr. 2021. Cit. on p. 15.

HALKO, Nathan; MARTINSSON, Per-Gunnar; TROPP, Joel A. Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions. California Institute of Technology, 2009. Cit. on p. 22.

HANSON, Richard J. A numerical method for solving Fredholm integral equations of the first kind using singular values. **SIAM Journal on Numerical Analysis**, SIAM, v. 8, n. 3, p. 616–622, 1971. Cit. on p. 22.

HE, Sheng; LEANSE, Leon G.; FENG, Yanfang. Artificial intelligence and machine learning assisted drug delivery for effective treatment of infectious diseases. **Advanced Drug Delivery Reviews**, v. 178, p. 113922, Nov. 2021. ISSN 0169409X. DOI: 10.1016/j.addr.2021.113922. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0169409X2100315X>. Visited on: 22 Jan. 2022. Cit. on p. 15.

HIRAYAMA, HIROKI; XU, X; PANG, K SANDY. Viability of the vascularly perfused, recirculating rat intestine and intestine-liver preparations. **American Journal of Physiology-Gastrointestinal and Liver Physiology**, American Physiological Society Bethesda, MD, v. 257, n. 2, g249–g258, 1989. Cit. on p. 21.

HOUARI, Rima et al. Dimensionality reduction in data mining: A Copula approach. **Expert Systems with Applications**, Elsevier, v. 64, p. 247–260, 2016. Cit. on p. 15.

JAMES, Gareth et al. **An introduction to statistical learning**. [S.l.]: Springer, 2013. v. 112. Cit. on p. 23.

JOLLIFFE, I. T. **Principal Component Analysis**. New York: Springer-Verlag, 2002. (Springer Series in Statistics). Cit. on p. 22.

JOUBERT, Ruan et al. In vitro oral drug permeation models: the importance of taking physiological and physico-chemical factors into consideration. **Expert opinion on drug delivery**, Taylor & Francis, v. 14, n. 2, p. 179–187, 2017. Cit. on p. 14.

KAK, Subhash. The intrinsic dimensionality of data. **Circuits, Systems, and Signal Processing**, Springer, v. 40, n. 5, p. 2599–2607, 2021. Cit. on p. 15.

KAMIYA, Yusuke et al. Prediction of permeability across intestinal cell monolayers for 219 disparate chemicals using in vitro experimental coefficients in a pH gradient system and in silico analyses by trivariate linear regressions and machine learning. **Biochemical Pharmacology**, v. 192, p. 114749, Oct. 2021. ISSN 00062952. DOI: 10.1016/j.bcp.2021.114749. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0006295221003658>. Visited on: 23 Jan. 2022. Cit. on p. 16.

KOCH, Gregory; ZEMEL, Richard; SALAKHUTDINOV, Ruslan, et al. Siamese neural networks for one-shot image recognition. In: LILLE. ICML deep learning workshop. [S.l.: s.n.], 2015. v. 2, p. 0. Cit. on p. 23.

KRAMER, Oliver. K-Nearest Neighbors. In_____. **Dimensionality Reduction with Unsupervised Nearest Neighbors**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. v. 51, p. 13–23. Series Title: Intelligent Systems Reference Library. ISBN 978-3-642-38651-0 978-3-642-38652-7. DOI: 10.1007/978-3-642-38652-7_2. Available from: <http://link.springer.com/10.1007/978-3-642-38652-7_2>. Visited on: 17 Feb. 2022. Cit. on p. 23.

KUMAR, Rajnish et al. Prediction of Human Intestinal Absorption of Compounds Using Artificial Intelligence Techniques. **Current Drug Discovery Technologies**, v. 14, n. 4, 27 Oct. 2017. ISSN 15701638. DOI: 10.2174/1570163814666170404160911. Available from: <http://www.eurekaselect.com/151330/article>. Visited on: 23 Jan. 2022. Cit. on p. 15.

LANDRUM, Greg. **RDKit: Open-source cheminformatics**. [S.l.: s.n.], 2020. DOI: 10.5281/zenodo.3732262. Available from: <http://www.rdkit.org>. Cit. on p. 26.

LEE, Ming-Han et al. In Silico Prediction of Intestinal Permeability by Hierarchical Support Vector Regression. **International Journal of Molecular Sciences**, v. 21, n. 10, p. 3582, 19 May 2020. ISSN 1422-0067. DOI: 10.3390/ijms21103582. Available from: <https://www.mdpi.com/1422-0067/21/10/3582>. Visited on: 23 Jan. 2022. Cit. on p. 16.

LI, Yinhui et al. An efficient intrusion detection system based on support vector machines and gradually feature removal method. **Expert Systems with Applications**, v. 39, n. 1, p. 424–430, 2012. Cit. on p. 21.

MARTINEZ, Marilyn N; AMIDON, Gordon L. A mechanistic approach to understanding the factors affecting drug absorption: a review of fundamentals. **The Journal of Clinical Pharmacology**, Wiley Online Library, v. 42, n. 6, p. 620–643, 2002. Cit. on p. 19.

MCINNES, Leland; HEALY, John; MELVILLE, James. Umap: Uniform manifold approximation and projection for dimension reduction. **arXiv preprint arXiv:1802.03426**, 2018. Cit. on p. 22.

NALLAPERUMA, Dinithi et al. Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management. **IEEE Transactions on Intelligent Transportation Systems**, v. 20, n. 12, p. 4679–4690, Dec. 2019. ISSN 1524-9050, 1558-0016. DOI: 10.1109/TITS.2019.2924883. Available from: <https://ieeexplore.ieee.org/document/8759919/>. Visited on: 4 Apr. 2021. Cit. on p. 15.

NAYARISSERI, Anuraj et al. Artificial Intelligence, Big Data and Machine Learning Approaches in Precision Medicine & Drug Discovery. **Current Drug Targets**, v. 22, n. 6, p. 631–655, 23 Apr. 2021. ISSN 13894501. DOI: 10.2174/1389450122999210104205732. Available from: <https://www.eurekaselect.com/189908/article>. Visited on: 23 Jan. 2022. Cit. on p. 16.

NORIEGA, Leonardo. Multilayer perceptron tutorial. **School of Computing. Staffordshire University**, Citeseer, 2005. Cit. on p. 23.

ORME, M. Drug absorption in the gut. **BJA: British Journal of Anaesthesia**, Oxford University Press, v. 56, n. 1, p. 59–67, 1984. Cit. on p. 14.

PAL, Sankar K; MITRA, Sushmita. Multilayer perceptron, fuzzy sets, classifiaction, 1992. Cit. on p. 23.

PEREIRA, Rafael B. et al. Categorizing feature selection methods for multi-label classification. **Artificial Intelligence Review**, v. 49, n. 1, p. 57–78, Jan. 2018. Cit. on p. 21.

SCHÖLKOPF, Bernhard; SMOLA, Alexander; MÜLLER, Klaus-Robert. Nonlinear component analysis as a kernel eigenvalue problem. **Neural computation**, MIT Press, v. 10, n. 5, p. 1299–1319, 1998. Cit. on p. 22.

SHARMA, A; KUMAR, R; VARADWAJ, P. Prediction of mutagenicity of compounds by support vector machine. **Online J Bioinfo**, v. 12, n. 1, p. 9–17, 2011. Cit. on p. 16.

SHIN, Moonshik et al. Predicting the Absorption Potential of Chemical Compounds Through a Deep Learning Approach. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, v. 15, n. 2, p. 432–440, Mar. 2018. Conference Name: IEEE/ACM Transactions on Computational Biology and Bioinformatics. ISSN 1557-9964. DOI: 10.1109/TCBB.2016.2535233. Cit. on p. 14.

SINGH, K. P.; GUPTA, S.; BASANT, N. In silico prediction of cellular permeability of diverse chemicals using qualitative and quantitative SAR modeling approaches. **Chemometrics Intell Laboratory Syst.**, v. 140, p. 61–72, 2015. Cit. on p. 15.

SNOEK, Jasper; LAROCHELLE, Hugo; ADAMS, Ryan P. Practical bayesian optimization of machine learning algorithms. **Advances in neural information processing systems**, v. 25, 2012. Cit. on p. 24.

STORPIRTIS, S.; AL., et. **Farmacocinética Básica e Aplicada**. [S.l.]: GEN, 2011. v. 1. Cit. on p. 19.

STUCKMAN, Jeffrey; WALDEN, James; SCANDARIATO, Riccardo. The effect of dimensionality reduction on software vulnerability prediction models. **IEEE Transactions on Reliability**, IEEE, v. 66, n. 1, p. 17–37, 2016. Cit. on p. 16.

SZUBERT, Benjamin et al. Structure-preserving visualisation of high dimensional single-cell datasets. **Scientific reports**, Nature Publishing Group, v. 9, n. 1, p. 1–10, 2019. Cit. on p. 23.

TA, Giang Huong et al. Development of a Hierarchical Support Vector Regression-Based In Silico Model for Caco-2 Permeability. **Pharmaceutics**, v. 13, n. 2, p. 174, 28 Jan. 2021. ISSN 1999-4923. DOI: 10.3390/pharmaceutics13020174. Available from: <https://www.mdpi.com/1999-4923/13/2/174>. Visited on: 8 Feb. 2022. Cit. on pp. 14, 16.

TOZER, T. N.; ROWLAN, M. **Essentials of Pharmacokinetics and Pharmacodynamics**. [S.l.]: Wolters Kluwer, 2016. v. 2. Cit. on pp. 14, 19.

VERTZONI, Maria et al. Impact of regional differences along the gastrointestinal tract of healthy adults on oral drug absorption: An UNGAP review. **European Journal of Pharmaceutical Sciences**, v. 134, p. 153–175, June 2019. ISSN 09280987. DOI: 10.1016/j.ejps.2019.04.013. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0928098719301514>. Visited on: 16 Feb. 2022. Cit. on pp. 14, 20.

WAN, Youchuan; WANG, Mingwei, et al. A Feature Selection Method Based on Modified Binary Coded Ant Colony Optimization Algorithm. **Appl. Soft Comput.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 49, n. 100, p. 248–258, Dec. 2016. ISSN 1568-4946. Cit. on p. 21.

WAN, Zhi; XU, Yading; ŠAVIJA, Branko. On the use of machine learning models for prediction of compressive strength of concrete: Influence of dimensionality reduction on the model performance. **Materials**, Multidisciplinary Digital Publishing Institute, v. 14, n. 4, p. 713, 2021. Cit. on p. 16.

WANG, N.; AL, et. Predicting human intestinal absorption with modified random forest approach: a comprehensive evaluation of molecular representation, unbalanced data, and applicability domain issues. **RSC Advances**, v. 7, p. 19007–19018, 2016. DOI: 10.1039/c6ra28442f. Cit. on p. 25.

WANG, Wei; YE, Zhuyifan, et al. Computational pharmaceutics - A new paradigm of drug delivery. **Journal of Controlled Release**, v. 338, p. 119–136, Oct. 2021. ISSN 01683659. DOI: 10.1016/j.jconrel.2021.08.030. Available from:

<https://linkinghub.elsevier.com/retrieve/pii/S0168365921004363>. Visited on: 22 Jan. 2022. Cit. on p. 15.

YANG, Ming et al. A novel adaptive ensemble classification framework for ADME prediction. **RSC Advances**, v. 8, n. 21, p. 11661–11683, 2018. ISSN 2046-2069. DOI: 10.1039/C8RA01206G. Available from: <http://xlink.rsc.org/?DOI=C8RA01206G>. Visited on: 29 Jan. 2022. Cit. on p. 16.

YOUHANNA, Sonia; LAUSCHKE, Volker M. The Past, Present and Future of Intestinal In Vitro Cell Systems for Drug Absorption Studies. **Journal of Pharmaceutical Sciences**, v. 110, n. 1, p. 50–65, Jan. 2021. ISSN 00223549. DOI: 10.1016/j.xphs.2020.07.001. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0022354920303622>. Visited on: 22 Jan. 2022. Cit. on pp. 14, 19.

Appendix

APPENDIX

A

CROSS-VALIDATION SCORES

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
50	—	KNN	0.7406 ± 0.0170	0.6571 ± 0.1135
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.9177 ± 0.0126	0.7784 ± 0.0828
		RF	0.9990 ± 0.0022	0.8253 ± 0.0708
2	PCA	KNN	1.0000 ± 0.0000	0.7078 ± 0.0710
		MLP	0.7060 ± 0.0256	0.6931 ± 0.0793
		SVM	0.7982 ± 0.0175	0.6842 ± 0.0675
		RF	0.7867 ± 0.0182	0.7455 ± 0.0695
	TSVD	KNN	1.0000 ± 0.0000	0.6597 ± 0.0912
		MLP	0.6640 ± 0.0190	0.6563 ± 0.0759
		SVM	0.8249 ± 0.0309	0.6457 ± 0.0976
		RF	1.0000 ± 0.0000	0.6794 ± 0.0521
	KPCA	KNN	0.7479 ± 0.0144	0.7357 ± 0.0748
		MLP	0.7055 ± 0.0144	0.6896 ± 0.0979
		SVM	0.7097 ± 0.0148	0.6991 ± 0.0975
		RF	0.9995 ± 0.0017	0.7080 ± 0.0703
	UMAP	KNN	1.0000 ± 0.0000	0.7608 ± 0.0896
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.7086 ± 0.0767	0.6567 ± 0.1243

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		RF	1.0000 ± 0.0000	0.7742 ± 0.0629
	Ivis	KNN	1.0000 ± 0.0000	0.8310 ± 0.0684
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.9906 ± 0.0162	0.8260 ± 0.0909
		RF	1.0000 ± 0.0000	0.8210 ± 0.0762
3	PCA	KNN	0.7684 ± 0.0159	0.7314 ± 0.0812
		MLP	0.7427 ± 0.0116	0.7409 ± 0.0795
		SVM	0.9287 ± 0.0121	0.7022 ± 0.0886
		RF	0.9817 ± 0.0090	0.7736 ± 0.0698
	TSVD	KNN	1.0000 ± 0.0000	0.7268 ± 0.0813
		MLP	0.7128 ± 0.0262	0.6989 ± 0.0761
		SVM	0.7070 ± 0.0255	0.7128 ± 0.0724
		RF	0.8496 ± 0.0220	0.7177 ± 0.0857
	KPCA	KNN	1.0000 ± 0.0000	0.7366 ± 0.0892
		MLP	0.7516 ± 0.0192	0.7221 ± 0.1040
		SVM	0.7269 ± 0.0122	0.7275 ± 0.1194
		RF	1.0000 ± 0.0000	0.7268 ± 0.1054
	UMAP	KNN	1.0000 ± 0.0000	0.7697 ± 0.0598
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.6021 ± 0.2097	0.4991 ± 0.0484
		RF	1.0000 ± 0.0000	0.6894 ± 0.0739
	Ivis	KNN	0.9466 ± 0.0212	0.8219 ± 0.0753
		MLP	0.9906 ± 0.0095	0.8403 ± 0.0692
		SVM	0.9911 ± 0.0082	0.8545 ± 0.0765
		RF	0.9995 ± 0.0017	0.6842 ± 0.0576
4	PCA	KNN	0.8595 ± 0.0143	0.7416 ± 0.0983
		MLP	0.7511 ± 0.0172	0.7457 ± 0.0842
		SVM	0.5242 ± 0.0456	0.4991 ± 0.0484
		RF	1.0000 ± 0.0000	0.7736 ± 0.0568
	TSVD	KNN	0.8333 ± 0.0240	0.6946 ± 0.1162
		MLP	0.7558 ± 0.0179	0.7314 ± 0.0823

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		SVM	0.7364 ± 0.0163	0.7314 ± 0.0823
		RF	0.9990 ± 0.0022	0.7457 ± 0.0725
	KPCA	KNN	1.0000 ± 0.0000	0.7699 ± 0.0914
		MLP	0.7411 ± 0.0113	0.7318 ± 0.0988
		SVM	0.5405 ± 0.0811	0.4946 ± 0.0844
		RF	0.9403 ± 0.0167	0.7074 ± 0.1191
	UMAP	KNN	1.0000 ± 0.0000	0.6848 ± 0.0819
		MLP	0.9838 ± 0.0229	0.7697 ± 0.0598
		SVM	1.0000 ± 0.0000	0.6610 ± 0.0965
		RF	0.9565 ± 0.0092	0.7268 ± 0.1007
	Ivis	KNN	1.0000 ± 0.0000	0.7885 ± 0.0639
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.9890 ± 0.0072	0.8167 ± 0.0698
		RF	0.8732 ± 0.0300	0.6613 ± 0.1225
5	PCA	KNN	0.7055 ± 0.0176	0.6751 ± 0.0749
		MLP	0.7573 ± 0.0153	0.7457 ± 0.0759
		SVM	0.9963 ± 0.0035	0.7361 ± 0.0582
		RF	1.0000 ± 0.0000	0.8115 ± 0.0483
	TSVD	KNN	0.7453 ± 0.0188	0.6939 ± 0.0716
		MLP	0.7463 ± 0.0088	0.7316 ± 0.0808
		SVM	0.7406 ± 0.0138	0.7219 ± 0.0709
		RF	0.8506 ± 0.0141	0.7314 ± 0.1135
	KPCA	KNN	1.0000 ± 0.0000	0.7359 ± 0.0808
		MLP	0.7165 ± 0.0171	0.7273 ± 0.1027
		SVM	0.5426 ± 0.0843	0.5037 ± 0.0948
		RF	1.0000 ± 0.0000	0.8067 ± 0.0508
	UMAP	KNN	0.9292 ± 0.0105	0.7879 ± 0.0747
		MLP	1.0000 ± 0.0000	0.6991 ± 0.1009
		SVM	1.0000 ± 0.0000	0.7416 ± 0.0823
		RF	0.9251 ± 0.0118	0.7879 ± 0.0712
	Ivis	KNN	1.0000 ± 0.0000	0.7654 ± 0.1018

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		MLP	0.9837 ± 0.0153	0.8260 ± 0.0689
		SVM	0.9859 ± 0.0096	0.8260 ± 0.0727
		RF	0.9617 ± 0.0262	0.7978 ± 0.0903
6	PCA	KNN	1.0000 ± 0.0000	0.7593 ± 0.0726
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.9990 ± 0.0022	0.7266 ± 0.0877
		RF	1.0000 ± 0.0000	0.7738 ± 0.0755
	TSVD	KNN	1.0000 ± 0.0000	0.7461 ± 0.0799
		MLP	0.7767 ± 0.0144	0.7595 ± 0.0946
		SVM	0.9764 ± 0.0114	0.7314 ± 0.0900
		RF	1.0000 ± 0.0000	0.7639 ± 0.0930
	KPCA	KNN	1.0000 ± 0.0000	0.7881 ± 0.0847
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	0.5000 ± 0.0025	0.5000 ± 0.0224
	UMAP	KNN	1.0000 ± 0.0000	0.7275 ± 0.0893
		MLP	0.9900 ± 0.0072	0.7697 ± 0.0598
		SVM	0.5137 ± 0.0238	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.7690 ± 0.0642
	Ivis	KNN	1.0000 ± 0.0000	0.8301 ± 0.0560
		MLP	0.9492 ± 0.0318	0.7498 ± 0.1218
		SVM	0.9921 ± 0.0057	0.8022 ± 0.0650
		RF	1.0000 ± 0.0000	0.7558 ± 0.0998
7	PCA	KNN	1.0000 ± 0.0000	0.7457 ± 0.1006
		MLP	0.7626 ± 0.0141	0.7459 ± 0.1015
		SVM	1.0000 ± 0.0000	0.7262 ± 0.0541
		RF	1.0000 ± 0.0000	0.7874 ± 0.0520
	TSVD	KNN	1.0000 ± 0.0000	0.6989 ± 0.0855
		MLP	0.7568 ± 0.0183	0.7364 ± 0.0841
		SVM	0.7348 ± 0.0116	0.7173 ± 0.0872
		RF	1.0000 ± 0.0000	0.7686 ± 0.0913

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
	KPCA	KNN	0.7464 ± 0.0259	0.7035 ± 0.0587
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.7348 ± 0.0119	0.6987 ± 0.1130
		RF	1.0000 ± 0.0000	0.7922 ± 0.0606
	UMAP	KNN	1.0000 ± 0.0000	0.7604 ± 0.0918
		MLP	1.0000 ± 0.0000	0.6801 ± 0.0929
		SVM	0.7778 ± 0.0458	0.7227 ± 0.0821
		RF	1.0000 ± 0.0000	0.6141 ± 0.0919
	Ivis	KNN	0.9025 ± 0.0229	0.7838 ± 0.1003
		MLP	0.9864 ± 0.0121	0.8212 ± 0.0720
		SVM	0.8653 ± 0.0247	0.7697 ± 0.0893
		RF	0.8616 ± 0.0264	0.7600 ± 0.0855
8	PCA	KNN	1.0000 ± 0.0000	0.7452 ± 0.0601
		MLP	0.7699 ± 0.0144	0.7550 ± 0.0942
		SVM	0.9759 ± 0.0066	0.7312 ± 0.0658
		RF	1.0000 ± 0.0000	0.7924 ± 0.0710
	TSVD	KNN	1.0000 ± 0.0000	0.7879 ± 0.0623
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.7594 ± 0.0192	0.7219 ± 0.0709
		RF	0.8197 ± 0.0159	0.7359 ± 0.0985
	KPCA	KNN	1.0000 ± 0.0000	0.7173 ± 0.0759
		MLP	0.9885 ± 0.0089	0.7400 ± 0.0622
		SVM	0.7400 ± 0.0141	0.7312 ± 0.0705
		RF	1.0000 ± 0.0000	0.7550 ± 0.0722
	UMAP	KNN	0.8868 ± 0.0240	0.7740 ± 0.0640
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	1.0000 ± 0.0000	0.6565 ± 0.0877
		RF	0.9282 ± 0.0133	0.7883 ± 0.0572
	Ivis	KNN	1.0000 ± 0.0000	0.8258 ± 0.0488
		MLP	0.9969 ± 0.0037	0.8117 ± 0.0576
		SVM	0.9906 ± 0.0139	0.8260 ± 0.0937

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		RF	1.0000 ± 0.0000	0.7273 ± 0.1034
9	PCA	KNN	1.0000 ± 0.0000	0.7266 ± 0.0569
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.7620 ± 0.0226	0.7219 ± 0.0672
		RF	1.0000 ± 0.0000	0.7779 ± 0.0809
	TSVD	KNN	1.0000 ± 0.0000	0.7926 ± 0.0544
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	1.0000 ± 0.0000	0.7398 ± 0.0938
		RF	0.9995 ± 0.0017	0.7688 ± 0.0766
	KPCA	KNN	1.0000 ± 0.0000	0.6989 ± 0.0698
		MLP	0.9806 ± 0.0102	0.7214 ± 0.0902
		SVM	0.7600 ± 0.0175	0.7312 ± 0.0588
		RF	1.0000 ± 0.0000	0.7690 ± 0.0710
	UMAP	KNN	1.0000 ± 0.0000	0.6896 ± 0.0988
		MLP	1.0000 ± 0.0000	0.6610 ± 0.0883
		SVM	0.5526 ± 0.1054	0.5219 ± 0.1015
		RF	1.0000 ± 0.0000	0.6662 ± 0.0923
	Ivis	KNN	0.9466 ± 0.0248	0.8214 ± 0.0709
		MLP	0.9229 ± 0.0317	0.7978 ± 0.0835
		SVM	0.9188 ± 0.0208	0.7974 ± 0.0770
		RF	1.0000 ± 0.0000	0.8028 ± 0.0878
10	PCA	KNN	1.0000 ± 0.0000	0.7264 ± 0.0621
		MLP	0.5220 ± 0.0373	0.5333 ± 0.0826
		SVM	0.7594 ± 0.0210	0.7223 ± 0.0950
		RF	0.9958 ± 0.0048	0.8160 ± 0.0747
	TSVD	KNN	1.0000 ± 0.0000	0.7455 ± 0.0668
		MLP	0.9371 ± 0.0226	0.7266 ± 0.1026
		SVM	0.7605 ± 0.0245	0.7266 ± 0.0753
		RF	0.9953 ± 0.0039	0.7597 ± 0.0860
	KPCA	KNN	0.7091 ± 0.0198	0.6762 ± 0.1133
		MLP	0.9502 ± 0.0229	0.7403 ± 0.1008

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		SVM	0.7621 ± 0.0092	0.7177 ± 0.1267
		RF	0.9995 ± 0.0017	0.7595 ± 0.0848
	UMAP	KNN	1.0000 ± 0.0000	0.6610 ± 0.0965
		MLP	1.0000 ± 0.0000	0.6610 ± 0.0965
		SVM	0.5079 ± 0.0141	0.4857 ± 0.0166
		RF	0.9334 ± 0.0124	0.7835 ± 0.0476
	Ivis	KNN	0.8852 ± 0.0226	0.7935 ± 0.0887
		MLP	0.9287 ± 0.0198	0.7885 ± 0.0901
		SVM	0.9948 ± 0.0035	0.8212 ± 0.0846
		RF	1.0000 ± 0.0000	0.7976 ± 0.1088
11	PCA	KNN	1.0000 ± 0.0000	0.7314 ± 0.0620
		MLP	0.7945 ± 0.0216	0.7600 ± 0.1103
		SVM	0.7584 ± 0.0178	0.7314 ± 0.0812
		RF	0.9534 ± 0.0149	0.7974 ± 0.0689
	TSVD	KNN	1.0000 ± 0.0000	0.7409 ± 0.0487
		MLP	0.9948 ± 0.0060	0.7305 ± 0.0885
		SVM	0.7610 ± 0.0231	0.7312 ± 0.0805
		RF	0.9995 ± 0.0017	0.7738 ± 0.0708
	KPCA	KNN	1.0000 ± 0.0000	0.7368 ± 0.0763
		MLP	0.9937 ± 0.0041	0.7361 ± 0.0537
		SVM	0.7568 ± 0.0207	0.7219 ± 0.0765
		RF	1.0000 ± 0.0000	0.8158 ± 0.0790
	UMAP	KNN	0.9916 ± 0.0075	0.7515 ± 0.0902
		MLP	1.0000 ± 0.0000	0.6610 ± 0.0883
		SVM	1.0000 ± 0.0000	0.7087 ± 0.0943
		RF	1.0000 ± 0.0000	0.7323 ± 0.0897
	Ivis	KNN	1.0000 ± 0.0000	0.7747 ± 0.0746
		MLP	0.7406 ± 0.0193	0.7175 ± 0.0975
		SVM	0.9528 ± 0.0231	0.8446 ± 0.0628
		RF	0.9607 ± 0.0151	0.7604 ± 0.1195
12	PCA	KNN	0.7065 ± 0.0162	0.6799 ± 0.0899

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		MLP	0.9874 ± 0.0096	0.7693 ± 0.1026
		SVM	0.7600 ± 0.0160	0.7409 ± 0.0983
		RF	1.0000 ± 0.0000	0.7922 ± 0.0710
	TSVD	KNN	1.0000 ± 0.0000	0.7409 ± 0.0536
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.5384 ± 0.0755	0.4946 ± 0.0504
		RF	0.9701 ± 0.0116	0.7829 ± 0.0831
	KPCA	KNN	1.0000 ± 0.0000	0.7177 ± 0.0886
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	1.0000 ± 0.0000	0.6463 ± 0.0952
		RF	1.0000 ± 0.0000	0.7920 ± 0.0647
	UMAP	KNN	1.0000 ± 0.0000	0.7738 ± 0.0654
		MLP	1.0000 ± 0.0000	0.7561 ± 0.0820
		SVM	0.9602 ± 0.0131	0.7838 ± 0.0838
		RF	0.9114 ± 0.0136	0.7788 ± 0.0564
	Ivis	KNN	1.0000 ± 0.0000	0.7974 ± 0.0852
		MLP	0.8172 ± 0.1613	0.7216 ± 0.1318
		SVM	1.0000 ± 0.0000	0.6898 ± 0.0995
		RF	1.0000 ± 0.0000	0.8307 ± 0.0765
13	PCA	KNN	0.8097 ± 0.0190	0.7314 ± 0.0427
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	1.0000 ± 0.0000	0.6275 ± 0.1079
		RF	1.0000 ± 0.0000	0.8065 ± 0.0523
	TSVD	KNN	1.0000 ± 0.0000	0.6991 ± 0.0901
		MLP	0.9051 ± 0.0210	0.7786 ± 0.0749
		SVM	1.0000 ± 0.0000	0.7264 ± 0.0698
		RF	0.9015 ± 0.0210	0.7881 ± 0.0828
	KPCA	KNN	0.6604 ± 0.0220	0.6846 ± 0.0796
		MLP	0.7647 ± 0.0116	0.7368 ± 0.1139
		SVM	0.7594 ± 0.0185	0.7316 ± 0.0951
		RF	0.9995 ± 0.0017	0.7180 ± 0.0888

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
	UMAP	KNN	0.9859 ± 0.0025	0.7558 ± 0.0805
		MLP	0.9654 ± 0.0187	0.7271 ± 0.0582
		SVM	0.9984 ± 0.0036	0.7459 ± 0.0562
		RF	1.0000 ± 0.0000	0.6848 ± 0.0934
	Ivis	KNN	1.0000 ± 0.0000	0.8071 ± 0.0702
		MLP	0.9995 ± 0.0017	0.8258 ± 0.0700
		SVM	0.9953 ± 0.0039	0.8162 ± 0.0560
		RF	1.0000 ± 0.0000	0.8214 ± 0.0709
14	PCA	KNN	1.0000 ± 0.0000	0.7548 ± 0.0526
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	0.9838 ± 0.0084	0.7974 ± 0.0436
	TSVD	KNN	0.8003 ± 0.0179	0.7175 ± 0.1111
		MLP	0.5026 ± 0.0012	0.4810 ± 0.0100
		SVM	0.5974 ± 0.1997	0.4991 ± 0.0484
		RF	1.0000 ± 0.0000	0.8019 ± 0.0814
	KPCA	KNN	1.0000 ± 0.0000	0.6939 ± 0.0818
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	1.0000 ± 0.0000	0.6985 ± 0.0825
		RF	1.0000 ± 0.0000	0.7688 ± 0.0598
	UMAP	KNN	1.0000 ± 0.0000	0.7366 ± 0.0703
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.5010 ± 0.0756	0.4634 ± 0.1085
		RF	0.0000 ± 0.0000	0.0000 ± 0.0000
	Ivis	KNN	0.8727 ± 0.0385	0.7699 ± 0.1105
		MLP	0.9397 ± 0.0268	0.8165 ± 0.0741
		SVM	0.9906 ± 0.0081	0.8069 ± 0.0872
		RF	1.0000 ± 0.0000	0.7972 ± 0.0999
15	PCA	KNN	1.0000 ± 0.0000	0.7500 ± 0.0492
		MLP	0.9979 ± 0.0037	0.7833 ± 0.0531
		SVM	1.0000 ± 0.0000	0.7405 ± 0.0554

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		RF	1.0000 ± 0.0000	0.7922 ± 0.0465
	TSVD	KNN	1.0000 ± 0.0000	0.7504 ± 0.0821
		MLP	0.7662 ± 0.0181	0.7364 ± 0.1101
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.7877 ± 0.0832
	KPCA	KNN	1.0000 ± 0.0000	0.7468 ± 0.1036
		MLP	0.5257 ± 0.0200	0.4952 ± 0.0446
		SVM	0.7558 ± 0.0231	0.7128 ± 0.1153
		RF	1.0000 ± 0.0000	0.7970 ± 0.0454
	UMAP	KNN	1.0000 ± 0.0000	0.7883 ± 0.0522
		MLP	1.0000 ± 0.0000	0.7649 ± 0.0663
		SVM	0.8753 ± 0.0290	0.7645 ± 0.0569
		RF	1.0000 ± 0.0000	0.7511 ± 0.0846
	Ivis	KNN	1.0000 ± 0.0000	0.7082 ± 0.0828
		MLP	0.9654 ± 0.0168	0.8262 ± 0.0710
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.8303 ± 0.0747
16	PCA	KNN	1.0000 ± 0.0000	0.7277 ± 0.0932
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.8162 ± 0.0451
	TSVD	KNN	0.7993 ± 0.0171	0.7173 ± 0.0812
		MLP	0.9790 ± 0.0105	0.7740 ± 0.1163
		SVM	0.5079 ± 0.0157	0.4855 ± 0.0224
		RF	0.9974 ± 0.0028	0.8160 ± 0.0916
	KPCA	KNN	1.0000 ± 0.0000	0.7411 ± 0.0687
		MLP	0.6258 ± 0.0818	0.5569 ± 0.1007
		SVM	0.7883 ± 0.0300	0.7504 ± 0.1096
		RF	0.5000 ± 0.0025	0.5000 ± 0.0224
	UMAP	KNN	1.0000 ± 0.0000	0.7835 ± 0.0572
		MLP	0.9948 ± 0.0074	0.7463 ± 0.0758

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		SVM	1.0000 ± 0.0000	0.7509 ± 0.0690
		RF	1.0000 ± 0.0000	0.6329 ± 0.0836
	Ivis	KNN	1.0000 ± 0.0000	0.7792 ± 0.0934
		MLP	0.9067 ± 0.0306	0.7699 ± 0.1160
		SVM	0.9591 ± 0.0266	0.8262 ± 0.0751
		RF	0.9937 ± 0.0065	0.7881 ± 0.0833
17	PCA	KNN	1.0000 ± 0.0000	0.7271 ± 0.0856
		MLP	0.9984 ± 0.0025	0.7552 ± 0.0506
		SVM	1.0000 ± 0.0000	0.7173 ± 0.0955
		RF	0.9597 ± 0.0126	0.7790 ± 0.0705
	TSVD	KNN	1.0000 ± 0.0000	0.7225 ± 0.0896
		MLP	0.9979 ± 0.0027	0.7552 ± 0.0774
		SVM	0.7631 ± 0.0158	0.7227 ± 0.0880
		RF	0.9911 ± 0.0055	0.7978 ± 0.0962
	KPCA	KNN	1.0000 ± 0.0000	0.7561 ± 0.1108
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.7542 ± 0.0279	0.7177 ± 0.1247
		RF	0.9486 ± 0.0158	0.7929 ± 0.0572
	UMAP	KNN	1.0000 ± 0.0000	0.6846 ± 0.0891
		MLP	0.5288 ± 0.0883	0.5143 ± 0.0729
		SVM	1.0000 ± 0.0000	0.6989 ± 0.0944
		RF	0.9969 ± 0.0044	0.7835 ± 0.0727
	Ivis	KNN	0.9534 ± 0.0151	0.8022 ± 0.0824
		MLP	0.9990 ± 0.0022	0.8115 ± 0.0798
		SVM	0.9481 ± 0.0239	0.8024 ± 0.0680
		RF	1.0000 ± 0.0000	0.8212 ± 0.0816
18	PCA	KNN	0.6672 ± 0.0260	0.6422 ± 0.0825
		MLP	0.7610 ± 0.0129	0.7128 ± 0.1014
		SVM	0.7605 ± 0.0240	0.7364 ± 0.0953
		RF	0.9151 ± 0.0204	0.7693 ± 0.0738
	TSVD	KNN	0.7327 ± 0.0266	0.6801 ± 0.0811

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		MLP	0.5516 ± 0.1033	0.4900 ± 0.0706
		SVM	0.7767 ± 0.0231	0.7221 ± 0.0793
		RF	1.0000 ± 0.0000	0.7840 ± 0.0943
	KPCA	KNN	1.0000 ± 0.0000	0.7641 ± 0.0500
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	1.0000 ± 0.0000	0.7128 ± 0.0567
		RF	0.8962 ± 0.0219	0.7597 ± 0.0632
	UMAP	KNN	1.0000 ± 0.0000	0.7325 ± 0.1088
		MLP	0.9989 ± 0.0033	0.7411 ± 0.0521
		SVM	0.9392 ± 0.0121	0.7459 ± 0.0983
		RF	1.0000 ± 0.0000	0.8022 ± 0.0521
	Ivis	KNN	1.0000 ± 0.0000	0.7933 ± 0.1024
		MLP	0.9806 ± 0.0116	0.8212 ± 0.0931
		SVM	0.9345 ± 0.0295	0.8117 ± 0.0658
		RF	1.0000 ± 0.0000	0.8167 ± 0.0860
19	PCA	KNN	0.7170 ± 0.0331	0.6424 ± 0.0812
		MLP	0.7563 ± 0.0126	0.6985 ± 0.0756
		SVM	1.0000 ± 0.0000	0.6658 ± 0.1230
		RF	0.9565 ± 0.0102	0.7734 ± 0.0758
	TSVD	KNN	1.0000 ± 0.0000	0.7593 ± 0.0845
		MLP	0.9995 ± 0.0017	0.7316 ± 0.1099
		SVM	0.7788 ± 0.0262	0.7266 ± 0.0817
		RF	1.0000 ± 0.0000	0.7742 ± 0.1048
	KPCA	KNN	0.8653 ± 0.0237	0.7312 ± 0.0975
		MLP	0.7720 ± 0.0446	0.6892 ± 0.1121
		SVM	0.5589 ± 0.1191	0.5082 ± 0.0800
		RF	1.0000 ± 0.0000	0.7318 ± 0.0757
	UMAP	KNN	1.0000 ± 0.0000	0.7931 ± 0.0688
		MLP	0.9292 ± 0.0120	0.7835 ± 0.0526
		SVM	1.0000 ± 0.0000	0.6610 ± 0.0883
		RF	1.0000 ± 0.0000	0.6610 ± 0.0938

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
	Ivis	KNN	1.0000 ± 0.0000	0.7972 ± 0.0978
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.9282 ± 0.0213	0.7835 ± 0.0852
		RF	1.0000 ± 0.0000	0.8071 ± 0.0802
20	PCA	KNN	1.0000 ± 0.0000	0.7084 ± 0.0868
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	1.0000 ± 0.0000	0.7266 ± 0.1106
		RF	1.0000 ± 0.0000	0.7833 ± 0.0619
	TSVD	KNN	0.6887 ± 0.0187	0.6561 ± 0.0461
		MLP	0.8989 ± 0.0135	0.7933 ± 0.0911
		SVM	0.5116 ± 0.0190	0.4855 ± 0.0224
		RF	0.9995 ± 0.0017	0.7929 ± 0.0902
	KPCA	KNN	1.0000 ± 0.0000	0.6846 ± 0.0919
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.7830 ± 0.0367	0.7087 ± 0.1157
		RF	0.9916 ± 0.0051	0.7974 ± 0.0662
	UMAP	KNN	1.0000 ± 0.0000	0.6994 ± 0.1155
		MLP	1.0000 ± 0.0000	0.6896 ± 0.0851
		SVM	1.0000 ± 0.0000	0.6751 ± 0.1048
		RF	0.9418 ± 0.0080	0.7885 ± 0.0714
	Ivis	KNN	1.0000 ± 0.0000	0.7840 ± 0.0989
		MLP	0.9832 ± 0.0228	0.8258 ± 0.0624
		SVM	0.7992 ± 0.0292	0.7370 ± 0.0975
		RF	1.0000 ± 0.0000	0.8353 ± 0.0739
21	PCA	KNN	0.8522 ± 0.0129	0.7084 ± 0.0714
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	1.0000 ± 0.0000	0.7032 ± 0.0685
		RF	0.8622 ± 0.0273	0.7933 ± 0.0948
	TSVD	KNN	1.0000 ± 0.0000	0.7223 ± 0.0683
		MLP	0.7726 ± 0.0227	0.7175 ± 0.0975
		SVM	0.7862 ± 0.0219	0.7271 ± 0.0875

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		RF	1.0000 ± 0.0000	0.7929 ± 0.1080
	KPCA	KNN	1.0000 ± 0.0000	0.7370 ± 0.0806
		MLP	0.8643 ± 0.0245	0.7552 ± 0.0670
		SVM	0.9418 ± 0.0130	0.8019 ± 0.0825
		RF	1.0000 ± 0.0000	0.8165 ± 0.1009
	UMAP	KNN	1.0000 ± 0.0000	0.6708 ± 0.0839
		MLP	0.8500 ± 0.2415	0.6758 ± 0.1432
		SVM	0.5153 ± 0.0267	0.4855 ± 0.0224
		RF	1.0000 ± 0.0000	0.7838 ± 0.0749
	Ivis	KNN	1.0000 ± 0.0000	0.8167 ± 0.0692
		MLP	0.9843 ± 0.0092	0.8452 ± 0.0864
		SVM	0.7699 ± 0.0164	0.7556 ± 0.0998
		RF	0.9576 ± 0.0147	0.8123 ± 0.0857
22	PCA	KNN	1.0000 ± 0.0000	0.7223 ± 0.0959
		MLP	0.9974 ± 0.0037	0.7543 ± 0.0454
		SVM	1.0000 ± 0.0000	0.6372 ± 0.0841
		RF	0.9974 ± 0.0037	0.8115 ± 0.0483
	TSVD	KNN	0.8548 ± 0.0197	0.7509 ± 0.0620
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.7940 ± 0.0287	0.7271 ± 0.0856
		RF	0.8690 ± 0.0232	0.7606 ± 0.1211
	KPCA	KNN	1.0000 ± 0.0000	0.7833 ± 0.0696
		MLP	0.5319 ± 0.0946	0.5000 ± 0.0594
		SVM	0.8061 ± 0.0228	0.7413 ± 0.1023
		RF	0.9209 ± 0.0186	0.7931 ± 0.0647
	UMAP	KNN	1.0000 ± 0.0000	0.7976 ± 0.0614
		MLP	0.9895 ± 0.0065	0.7606 ± 0.0792
		SVM	0.5026 ± 0.0012	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.6706 ± 0.0965
	Ivis	KNN	0.8899 ± 0.0327	0.7978 ± 0.0893
		MLP	0.9607 ± 0.0223	0.8258 ± 0.0663

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		SVM	0.9990 ± 0.0022	0.8171 ± 0.0924
		RF	1.0000 ± 0.0000	0.8165 ± 0.0925
23	PCA	KNN	0.8569 ± 0.0142	0.7080 ± 0.0738
		MLP	0.8863 ± 0.0172	0.7509 ± 0.1055
		SVM	0.7919 ± 0.0231	0.7266 ± 0.0569
		RF	1.0000 ± 0.0000	0.7831 ± 0.0593
	TSVD	KNN	1.0000 ± 0.0000	0.6755 ± 0.1012
		MLP	0.9984 ± 0.0035	0.7786 ± 0.0486
		SVM	0.7982 ± 0.0281	0.7318 ± 0.0800
		RF	0.9534 ± 0.0134	0.7933 ± 0.0911
	KPCA	KNN	0.8648 ± 0.0165	0.7361 ± 0.0663
		MLP	0.7301 ± 0.0859	0.6799 ± 0.1131
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.7784 ± 0.0544
	UMAP	KNN	0.9937 ± 0.0081	0.7087 ± 0.1018
		MLP	1.0000 ± 0.0000	0.6517 ± 0.0939
		SVM	1.0000 ± 0.0000	0.7595 ± 0.0677
		RF	1.0000 ± 0.0000	0.7273 ± 0.0816
	Ivis	KNN	1.0000 ± 0.0000	0.8167 ± 0.0692
		MLP	0.9649 ± 0.0157	0.8258 ± 0.0700
		SVM	0.9995 ± 0.0017	0.7976 ± 0.1129
		RF	1.0000 ± 0.0000	0.7835 ± 0.0724
24	PCA	KNN	0.7049 ± 0.0208	0.6281 ± 0.0800
		MLP	0.7720 ± 0.0148	0.7134 ± 0.1252
		SVM	1.0000 ± 0.0000	0.7268 ± 0.0841
		RF	0.8753 ± 0.0228	0.7838 ± 0.0776
	TSVD	KNN	0.8150 ± 0.0229	0.7459 ± 0.1120
		MLP	0.9565 ± 0.0099	0.7788 ± 0.0942
		SVM	0.7930 ± 0.0220	0.7416 ± 0.0964
		RF	0.9749 ± 0.0095	0.7838 ± 0.1179
	KPCA	KNN	1.0000 ± 0.0000	0.7604 ± 0.0861

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.8422 ± 0.0204	0.7604 ± 0.1372
		RF	0.8742 ± 0.0252	0.7647 ± 0.0777
	UMAP	KNN	1.0000 ± 0.0000	0.7883 ± 0.0715
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	1.0000 ± 0.0000	0.7697 ± 0.0598
		RF	1.0000 ± 0.0000	0.6563 ± 0.0999
	Ivis	KNN	1.0000 ± 0.0000	0.7840 ± 0.0891
		MLP	0.9937 ± 0.0081	0.8258 ± 0.0700
		SVM	0.9932 ± 0.0078	0.8024 ± 0.0680
		RF	1.0000 ± 0.0000	0.8071 ± 0.0833
25	PCA	KNN	1.0000 ± 0.0000	0.6942 ± 0.0920
		MLP	0.7746 ± 0.0214	0.7039 ± 0.1097
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	0.9921 ± 0.0045	0.8067 ± 0.0606
	TSVD	KNN	1.0000 ± 0.0000	0.7227 ± 0.0789
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.8223 ± 0.0346	0.7128 ± 0.0724
		RF	0.9921 ± 0.0062	0.7699 ± 0.0993
	KPCA	KNN	1.0000 ± 0.0000	0.7701 ± 0.0951
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.8412 ± 0.0247	0.7455 ± 0.0827
		RF	1.0000 ± 0.0000	0.7976 ± 0.0808
	UMAP	KNN	1.0000 ± 0.0000	0.7838 ± 0.0640
		MLP	0.9906 ± 0.0069	0.7697 ± 0.0598
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.7883 ± 0.0614
	Ivis	KNN	1.0000 ± 0.0000	0.8210 ± 0.0616
		MLP	0.9911 ± 0.0151	0.7974 ± 0.0767
		SVM	1.0000 ± 0.0000	0.7784 ± 0.0773

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		RF	1.0000 ± 0.0000	0.7597 ± 0.0591
26	PCA	KNN	0.8050 ± 0.0207	0.7457 ± 0.1175
		MLP	0.8743 ± 0.0852	0.7357 ± 0.0928
		SVM	0.7982 ± 0.0356	0.7318 ± 0.0890
		RF	0.9659 ± 0.0083	0.7976 ± 0.0883
	TSVD	KNN	1.0000 ± 0.0000	0.7364 ± 0.0521
		MLP	0.5393 ± 0.0444	0.5238 ± 0.0891
		SVM	1.0000 ± 0.0000	0.7690 ± 0.0840
		RF	0.9623 ± 0.0048	0.7887 ± 0.1143
	KPCA	KNN	1.0000 ± 0.0000	0.7558 ± 0.0770
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.8627 ± 0.0194	0.7742 ± 0.1012
		RF	1.0000 ± 0.0000	0.8026 ± 0.1251
	UMAP	KNN	1.0000 ± 0.0000	0.7933 ± 0.0810
		MLP	0.8166 ± 0.0351	0.7266 ± 0.0932
		SVM	0.0000 ± 0.0000	0.0000 ± 0.0000
		RF	1.0000 ± 0.0000	0.6517 ± 0.0939
	Ivis	KNN	1.0000 ± 0.0000	0.8076 ± 0.0676
		MLP	0.8994 ± 0.0269	0.8022 ± 0.0824
		SVM	0.9958 ± 0.0054	0.7976 ± 0.0691
		RF	1.0000 ± 0.0000	0.8307 ± 0.0939
27	PCA	KNN	1.0000 ± 0.0000	0.7461 ± 0.0733
		MLP	0.6515 ± 0.0661	0.6134 ± 0.0818
		SVM	0.7804 ± 0.0262	0.7037 ± 0.0759
		RF	1.0000 ± 0.0000	0.7786 ± 0.0761
	TSVD	KNN	1.0000 ± 0.0000	0.7277 ± 0.0904
		MLP	1.0000 ± 0.0000	0.7693 ± 0.0657
		SVM	0.8449 ± 0.0428	0.7225 ± 0.0889
		RF	0.9911 ± 0.0078	0.7929 ± 0.0799
	KPCA	KNN	0.7369 ± 0.0267	0.6797 ± 0.0752
		MLP	0.8899 ± 0.0143	0.7879 ± 0.0997

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		SVM	0.8438 ± 0.0259	0.7459 ± 0.0806
		RF	1.0000 ± 0.0000	0.8022 ± 0.0956
	UMAP	KNN	1.0000 ± 0.0000	0.7883 ± 0.0654
		MLP	0.9308 ± 0.0104	0.7883 ± 0.0614
		SVM	0.8926 ± 0.0247	0.7742 ± 0.0543
		RF	1.0000 ± 0.0000	0.6706 ± 0.0965
	Ivis	KNN	1.0000 ± 0.0000	0.7931 ± 0.0688
		MLP	0.8381 ± 0.0244	0.7697 ± 0.0872
		SVM	0.9340 ± 0.0248	0.7885 ± 0.1051
		RF	1.0000 ± 0.0000	0.7981 ± 0.1091
28	PCA	KNN	0.8669 ± 0.0128	0.7370 ± 0.0895
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.8721 ± 0.0336	0.7697 ± 0.0835
		RF	0.8989 ± 0.0292	0.7976 ± 0.0727
	TSVD	KNN	1.0000 ± 0.0000	0.7459 ± 0.0806
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.8585 ± 0.0358	0.7225 ± 0.0742
		RF	0.9990 ± 0.0022	0.8067 ± 0.0714
	KPCA	KNN	1.0000 ± 0.0000	0.7364 ± 0.1043
		MLP	0.8475 ± 0.0160	0.7786 ± 0.1032
		SVM	0.8679 ± 0.0200	0.7929 ± 0.0799
		RF	1.0000 ± 0.0000	0.7643 ± 0.0858
	UMAP	KNN	1.0000 ± 0.0000	0.6658 ± 0.0926
		MLP	1.0000 ± 0.0000	0.6658 ± 0.0870
		SVM	0.9282 ± 0.0102	0.7978 ± 0.0677
		RF	1.0000 ± 0.0000	0.6846 ± 0.0919
	Ivis	KNN	1.0000 ± 0.0000	0.8262 ± 0.0840
		MLP	0.9879 ± 0.0111	0.8167 ± 0.0884
		SVM	0.9969 ± 0.0037	0.8071 ± 0.0946
		RF	1.0000 ± 0.0000	0.8071 ± 0.0540
29	PCA	KNN	1.0000 ± 0.0000	0.7457 ± 0.0577

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		MLP	0.8569 ± 0.0233	0.7976 ± 0.0924
		SVM	1.0000 ± 0.0000	0.7502 ± 0.0916
		RF	1.0000 ± 0.0000	0.7887 ± 0.0829
	TSVD	KNN	1.0000 ± 0.0000	0.7699 ± 0.0886
		MLP	1.0000 ± 0.0000	0.7738 ± 0.0465
		SVM	1.0000 ± 0.0000	0.6227 ± 0.0623
		RF	1.0000 ± 0.0000	0.7883 ± 0.0939
	KPCA	KNN	0.8711 ± 0.0118	0.7792 ± 0.0622
		MLP	0.9796 ± 0.0072	0.7738 ± 0.1011
		SVM	0.8826 ± 0.0213	0.7970 ± 0.0981
		RF	1.0000 ± 0.0000	0.8115 ± 0.0798
	UMAP	KNN	1.0000 ± 0.0000	0.8119 ± 0.0684
		MLP	1.0000 ± 0.0000	0.6465 ± 0.0796
		SVM	0.9664 ± 0.0097	0.7788 ± 0.0517
		RF	1.0000 ± 0.0000	0.6565 ± 0.1011
	Ivis	KNN	0.9256 ± 0.0150	0.8026 ± 0.0920
		MLP	0.9921 ± 0.0083	0.8069 ± 0.0928
		SVM	0.9686 ± 0.0217	0.8162 ± 0.0871
		RF	1.0000 ± 0.0000	0.7838 ± 0.0708
30	PCA	KNN	1.0000 ± 0.0000	0.7742 ± 0.0773
		MLP	0.5796 ± 0.0702	0.5708 ± 0.1054
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.7976 ± 0.0614
	TSVD	KNN	0.7935 ± 0.0106	0.7366 ± 0.0593
		MLP	0.8412 ± 0.0196	0.7418 ± 0.1073
		SVM	0.8632 ± 0.0131	0.7887 ± 0.0916
		RF	0.9953 ± 0.0068	0.8117 ± 0.1010
	KPCA	KNN	1.0000 ± 0.0000	0.7649 ± 0.0768
		MLP	0.8595 ± 0.0147	0.7883 ± 0.0996
		SVM	0.8768 ± 0.0169	0.7885 ± 0.0977
		RF	0.9146 ± 0.0214	0.8115 ± 0.0541

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
	UMAP	KNN	1.0000 ± 0.0000	0.7933 ± 0.0710
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.7182 ± 0.1021
	Ivis	KNN	1.0000 ± 0.0000	0.8071 ± 0.0833
		MLP	0.9906 ± 0.0120	0.8028 ± 0.0912
		SVM	0.9334 ± 0.0145	0.7972 ± 0.1229
		RF	0.9974 ± 0.0037	0.7935 ± 0.1083
31	PCA	KNN	1.0000 ± 0.0000	0.7323 ± 0.0869
		MLP	0.8873 ± 0.0191	0.7929 ± 0.0929
		SVM	0.8506 ± 0.0168	0.8028 ± 0.0907
		RF	1.0000 ± 0.0000	0.8210 ± 0.0722
	TSVD	KNN	1.0000 ± 0.0000	0.7225 ± 0.0745
		MLP	0.9733 ± 0.0409	0.7697 ± 0.0931
		SVM	0.8821 ± 0.0131	0.7747 ± 0.1111
		RF	1.0000 ± 0.0000	0.8165 ± 0.0943
	KPCA	KNN	0.7086 ± 0.0278	0.6892 ± 0.0775
		MLP	0.8700 ± 0.0209	0.7974 ± 0.0910
		SVM	0.8857 ± 0.0194	0.7742 ± 0.0916
		RF	0.9219 ± 0.0159	0.7881 ± 0.0887
	UMAP	KNN	1.0000 ± 0.0000	0.7039 ± 0.1052
		MLP	0.9329 ± 0.0112	0.7881 ± 0.0577
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.6613 ± 0.0924
	Ivis	KNN	1.0000 ± 0.0000	0.8074 ± 0.0909
		MLP	0.9958 ± 0.0033	0.8165 ± 0.0549
		SVM	0.9183 ± 0.0167	0.8022 ± 0.0793
		RF	0.9769 ± 0.0119	0.8260 ± 0.0852
32	PCA	KNN	1.0000 ± 0.0000	0.7647 ± 0.0839
		MLP	0.8391 ± 0.0142	0.7881 ± 0.0915
		SVM	0.8957 ± 0.0121	0.7742 ± 0.0733

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		RF	1.0000 ± 0.0000	0.8162 ± 0.0603
	TSVD	KNN	0.7065 ± 0.0313	0.6424 ± 0.0843
		MLP	0.9995 ± 0.0017	0.7647 ± 0.0716
		SVM	0.8899 ± 0.0193	0.7792 ± 0.0986
		RF	0.9654 ± 0.0066	0.7790 ± 0.1100
	KPCA	KNN	1.0000 ± 0.0000	0.7506 ± 0.0708
		MLP	0.8700 ± 0.0168	0.7833 ± 0.0959
		SVM	0.8658 ± 0.0192	0.7931 ± 0.0755
		RF	0.9990 ± 0.0022	0.7835 ± 0.0640
	UMAP	KNN	1.0000 ± 0.0000	0.6753 ± 0.0922
		MLP	1.0000 ± 0.0000	0.6563 ± 0.0892
		SVM	0.9995 ± 0.0017	0.7502 ± 0.0488
		RF	0.9334 ± 0.0090	0.7740 ± 0.0460
	Ivis	KNN	1.0000 ± 0.0000	0.8123 ± 0.0993
		MLP	0.8905 ± 0.0133	0.7697 ± 0.1031
		SVM	0.9518 ± 0.0095	0.7879 ± 0.0871
		RF	0.9324 ± 0.0249	0.8026 ± 0.1095
33	PCA	KNN	1.0000 ± 0.0000	0.7511 ± 0.0835
		MLP	0.8947 ± 0.0072	0.8069 ± 0.0890
		SVM	0.8957 ± 0.0136	0.7883 ± 0.1045
		RF	1.0000 ± 0.0000	0.7976 ± 0.0525
	TSVD	KNN	1.0000 ± 0.0000	0.6950 ± 0.1175
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.8889 ± 0.0186	0.7887 ± 0.1038
		RF	0.9990 ± 0.0022	0.8019 ± 0.0692
	KPCA	KNN	0.7054 ± 0.0284	0.6569 ± 0.1103
		MLP	0.8506 ± 0.0175	0.7647 ± 0.0869
		SVM	0.8973 ± 0.0158	0.7693 ± 0.0895
		RF	0.9654 ± 0.0161	0.8305 ± 0.0799
	UMAP	KNN	1.0000 ± 0.0000	0.7697 ± 0.0598
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		SVM	0.9827 ± 0.0065	0.7697 ± 0.0717
		RF	1.0000 ± 0.0000	0.7788 ± 0.0758
	Ivis	KNN	1.0000 ± 0.0000	0.7879 ± 0.0809
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	1.0000 ± 0.0000	0.8450 ± 0.0849
		RF	1.0000 ± 0.0000	0.7784 ± 0.0668
34	PCA	KNN	0.8019 ± 0.0196	0.6610 ± 0.0653
		MLP	0.8889 ± 0.0087	0.8024 ± 0.0869
		SVM	0.9078 ± 0.0186	0.8024 ± 0.0995
		RF	0.9822 ± 0.0061	0.8028 ± 0.0824
	TSVD	KNN	1.0000 ± 0.0000	0.7318 ± 0.0663
		MLP	0.9995 ± 0.0017	0.8069 ± 0.0635
		SVM	0.8989 ± 0.0156	0.7742 ± 0.0943
		RF	0.9822 ± 0.0086	0.8121 ± 0.1019
	KPCA	KNN	1.0000 ± 0.0000	0.7794 ± 0.0752
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.8920 ± 0.0167	0.7788 ± 0.0871
		RF	0.5005 ± 0.0024	0.4952 ± 0.0219
	UMAP	KNN	0.9292 ± 0.0109	0.7838 ± 0.0678
		MLP	1.0000 ± 0.0000	0.6565 ± 0.1006
		SVM	1.0000 ± 0.0000	0.7039 ± 0.1003
		RF	1.0000 ± 0.0000	0.7554 ± 0.0488
	Ivis	KNN	1.0000 ± 0.0000	0.7879 ± 0.0387
		MLP	0.9775 ± 0.0210	0.7926 ± 0.0593
		SVM	0.9261 ± 0.0218	0.7879 ± 0.0777
		RF	1.0000 ± 0.0000	0.8167 ± 0.0727
35	PCA	KNN	0.6992 ± 0.0262	0.6615 ± 0.0788
		MLP	1.0000 ± 0.0000	0.7602 ± 0.0564
		SVM	0.9120 ± 0.0130	0.8026 ± 0.0940
		RF	1.0000 ± 0.0000	0.8303 ± 0.0710
	TSVD	KNN	0.7982 ± 0.0254	0.7554 ± 0.0579

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.8983 ± 0.0138	0.7792 ± 0.0886
		RF	0.9969 ± 0.0044	0.8165 ± 0.0710
	KPCA	KNN	1.0000 ± 0.0000	0.7364 ± 0.0724
		MLP	0.9052 ± 0.0140	0.8026 ± 0.0912
		SVM	0.9214 ± 0.0177	0.7558 ± 0.1065
		RF	1.0000 ± 0.0000	0.8351 ± 0.0778
	UMAP	KNN	1.0000 ± 0.0000	0.7740 ± 0.0508
		MLP	1.0000 ± 0.0000	0.6708 ± 0.0897
		SVM	0.9880 ± 0.0111	0.7643 ± 0.0491
		RF	1.0000 ± 0.0000	0.7509 ± 0.0797
	Ivis	KNN	1.0000 ± 0.0000	0.8398 ± 0.0502
		MLP	0.9502 ± 0.0211	0.7831 ± 0.0744
		SVM	0.9523 ± 0.0343	0.8167 ± 0.0825
		RF	1.0000 ± 0.0000	0.7929 ± 0.0860
36	PCA	KNN	0.7668 ± 0.0281	0.6797 ± 0.0746
		MLP	0.9020 ± 0.0081	0.8165 ± 0.0806
		SVM	0.8889 ± 0.0086	0.8026 ± 0.0794
		RF	0.9298 ± 0.0194	0.8119 ± 0.0606
	TSVD	KNN	0.8349 ± 0.0173	0.7457 ± 0.1019
		MLP	0.8910 ± 0.0139	0.7976 ± 0.0787
		SVM	0.9104 ± 0.0141	0.7885 ± 0.0838
		RF	0.9848 ± 0.0058	0.7929 ± 0.0692
	KPCA	KNN	0.7610 ± 0.0217	0.7136 ± 0.0965
		MLP	0.8774 ± 0.0182	0.7931 ± 0.1029
		SVM	0.9251 ± 0.0174	0.7604 ± 0.1061
		RF	1.0000 ± 0.0000	0.8305 ± 0.0543
	UMAP	KNN	1.0000 ± 0.0000	0.6472 ± 0.0954
		MLP	1.0000 ± 0.0000	0.6799 ± 0.0857
		SVM	0.9287 ± 0.0114	0.7931 ± 0.0688
		RF	0.9224 ± 0.0175	0.7790 ± 0.0675

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
	Ivis	KNN	0.7946 ± 0.0265	0.7656 ± 0.1009
		MLP	0.9764 ± 0.0185	0.8214 ± 0.0777
		SVM	0.9911 ± 0.0148	0.8165 ± 0.0837
		RF	0.9235 ± 0.0150	0.7747 ± 0.1187
37	PCA	KNN	1.0000 ± 0.0000	0.7693 ± 0.0439
		MLP	0.9041 ± 0.0112	0.8212 ± 0.0681
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.8212 ± 0.0708
	TSVD	KNN	1.0000 ± 0.0000	0.7738 ± 0.0419
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.9114 ± 0.0090	0.7976 ± 0.0868
		RF	1.0000 ± 0.0000	0.7935 ± 0.0829
	KPCA	KNN	1.0000 ± 0.0000	0.6991 ± 0.1009
		MLP	1.0000 ± 0.0000	0.7639 ± 0.0899
		SVM	0.9162 ± 0.0120	0.7929 ± 0.0767
		RF	1.0000 ± 0.0000	0.8212 ± 0.0512
	UMAP	KNN	1.0000 ± 0.0000	0.6610 ± 0.0965
		MLP	1.0000 ± 0.0000	0.6706 ± 0.0965
		SVM	1.0000 ± 0.0000	0.7561 ± 0.0820
		RF	1.0000 ± 0.0000	0.7323 ± 0.0869
	Ivis	KNN	1.0000 ± 0.0000	0.8169 ± 0.0687
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	1.0000 ± 0.0000	0.6325 ± 0.0743
		RF	1.0000 ± 0.0000	0.7792 ± 0.1084
38	PCA	KNN	1.0000 ± 0.0000	0.7216 ± 0.0653
		MLP	0.8952 ± 0.0106	0.8119 ± 0.0843
		SVM	0.9261 ± 0.0187	0.8022 ± 0.0813
		RF	0.9298 ± 0.0199	0.8165 ± 0.0670
	TSVD	KNN	1.0000 ± 0.0000	0.7080 ± 0.0738
		MLP	0.8978 ± 0.0145	0.7981 ± 0.0985
		SVM	0.9109 ± 0.0128	0.7931 ± 0.0979

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		RF	0.9979 ± 0.0027	0.7929 ± 0.0884
	KPCA	KNN	1.0000 ± 0.0000	0.7357 ± 0.0639
		MLP	0.8470 ± 0.0136	0.7597 ± 0.0943
		SVM	0.9272 ± 0.0168	0.7835 ± 0.0956
		RF	1.0000 ± 0.0000	0.8258 ± 0.0624
	UMAP	KNN	1.0000 ± 0.0000	0.6848 ± 0.0906
		MLP	1.0000 ± 0.0000	0.6284 ± 0.0929
		SVM	1.0000 ± 0.0000	0.7370 ± 0.0842
		RF	1.0000 ± 0.0000	0.8214 ± 0.0777
	Ivis	KNN	1.0000 ± 0.0000	0.7879 ± 0.0768
		MLP	0.9371 ± 0.0191	0.8305 ± 0.0629
		SVM	0.8999 ± 0.0102	0.7929 ± 0.0884
		RF	1.0000 ± 0.0000	0.8115 ± 0.0541
39	PCA	KNN	0.7332 ± 0.0181	0.6892 ± 0.0784
		MLP	0.9130 ± 0.0137	0.7976 ± 0.0720
		SVM	0.9057 ± 0.0152	0.7976 ± 0.0755
		RF	1.0000 ± 0.0000	0.8165 ± 0.0549
	TSVD	KNN	0.7532 ± 0.0296	0.6792 ± 0.0833
		MLP	0.9890 ± 0.0046	0.7926 ± 0.0589
		SVM	0.8973 ± 0.0113	0.7933 ± 0.0938
		RF	0.9214 ± 0.0174	0.7887 ± 0.0862
	KPCA	KNN	1.0000 ± 0.0000	0.7985 ± 0.1014
		MLP	0.9235 ± 0.0144	0.8022 ± 0.0724
		SVM	0.9036 ± 0.0162	0.7978 ± 0.0835
		RF	1.0000 ± 0.0000	0.7976 ± 0.0849
	UMAP	KNN	1.0000 ± 0.0000	0.7883 ± 0.0572
		MLP	1.0000 ± 0.0000	0.7229 ± 0.0927
		SVM	0.5753 ± 0.1532	0.5219 ± 0.0969
		RF	1.0000 ± 0.0000	0.6991 ± 0.0958
	Ivis	KNN	1.0000 ± 0.0000	0.8307 ± 0.0655
		MLP	0.9366 ± 0.0378	0.8167 ± 0.0889

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		SVM	0.9224 ± 0.0111	0.8310 ± 0.0901
		RF	1.0000 ± 0.0000	0.8258 ± 0.0700
40	PCA	KNN	1.0000 ± 0.0000	0.7130 ± 0.0921
		MLP	0.8066 ± 0.0161	0.7078 ± 0.0710
		SVM	0.9198 ± 0.0141	0.7929 ± 0.0794
		RF	1.0000 ± 0.0000	0.8212 ± 0.0819
	TSVD	KNN	0.8674 ± 0.0131	0.7126 ± 0.0894
		MLP	0.5016 ± 0.0018	0.4857 ± 0.0166
		SVM	0.9235 ± 0.0139	0.7745 ± 0.0982
		RF	1.0000 ± 0.0000	0.7978 ± 0.0875
	KPCA	KNN	1.0000 ± 0.0000	0.7693 ± 0.0856
		MLP	0.8522 ± 0.0153	0.7883 ± 0.0814
		SVM	0.9261 ± 0.0162	0.7602 ± 0.0897
		RF	0.5021 ± 0.0011	0.4810 ± 0.0100
	UMAP	KNN	1.0000 ± 0.0000	0.7933 ± 0.0840
		MLP	1.0000 ± 0.0000	0.6658 ± 0.0870
		SVM	1.0000 ± 0.0000	0.6706 ± 0.0965
		RF	1.0000 ± 0.0000	0.6468 ± 0.0641
	Ivis	KNN	1.0000 ± 0.0000	0.8024 ± 0.0642
		MLP	0.9434 ± 0.0155	0.7929 ± 0.0615
		SVM	0.9817 ± 0.0199	0.7922 ± 0.0821
		RF	1.0000 ± 0.0000	0.7835 ± 0.0679
41	PCA	KNN	0.7421 ± 0.0385	0.6554 ± 0.1122
		MLP	0.9004 ± 0.0165	0.8071 ± 0.0540
		SVM	0.9256 ± 0.0143	0.7695 ± 0.0763
		RF	0.9969 ± 0.0051	0.8214 ± 0.0673
	TSVD	KNN	1.0000 ± 0.0000	0.7175 ± 0.0675
		MLP	0.9120 ± 0.0147	0.7690 ± 0.0829
		SVM	0.9287 ± 0.0121	0.7740 ± 0.0805
		RF	0.9224 ± 0.0167	0.7931 ± 0.0790
	KPCA	KNN	0.5000 ± 0.0025	0.5000 ± 0.0224

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		MLP	0.9979 ± 0.0037	0.7929 ± 0.0727
		SVM	0.9272 ± 0.0151	0.7647 ± 0.0750
		RF	1.0000 ± 0.0000	0.8400 ± 0.0539
	UMAP	KNN	0.7820 ± 0.0583	0.7277 ± 0.0904
		MLP	1.0000 ± 0.0000	0.7325 ± 0.0965
		SVM	0.9208 ± 0.0128	0.7833 ± 0.0491
		RF	0.9313 ± 0.0107	0.7790 ± 0.0551
	Ivis	KNN	1.0000 ± 0.0000	0.8117 ± 0.0476
		MLP	0.9759 ± 0.0099	0.8260 ± 0.0651
		SVM	0.9644 ± 0.0197	0.8171 ± 0.1063
		RF	0.9633 ± 0.0092	0.8405 ± 0.0875
42	PCA	KNN	0.6557 ± 0.0302	0.6457 ± 0.0935
		MLP	0.9078 ± 0.0092	0.8165 ± 0.0670
		SVM	0.9162 ± 0.0107	0.8258 ± 0.0657
		RF	1.0000 ± 0.0000	0.8307 ± 0.0527
	TSVD	KNN	0.6383 ± 0.0274	0.5944 ± 0.0660
		MLP	0.8941 ± 0.0144	0.7840 ± 0.0860
		SVM	0.9125 ± 0.0148	0.7693 ± 0.0628
		RF	0.9796 ± 0.0068	0.8071 ± 0.0887
	KPCA	KNN	1.0000 ± 0.0000	0.7411 ± 0.0521
		MLP	0.9141 ± 0.0133	0.7833 ± 0.0797
		SVM	0.5589 ± 0.1187	0.5128 ± 0.0835
		RF	1.0000 ± 0.0000	0.8117 ± 0.0572
	UMAP	KNN	1.0000 ± 0.0000	0.7838 ± 0.0640
		MLP	0.9827 ± 0.0061	0.7461 ± 0.0628
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	0.9308 ± 0.0110	0.7788 ± 0.0522
	Ivis	KNN	1.0000 ± 0.0000	0.7786 ± 0.0622
		MLP	0.7746 ± 0.0298	0.7087 ± 0.0913
		SVM	0.5421 ± 0.0832	0.4946 ± 0.0504

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		RF	1.0000 ± 0.0000	0.8074 ± 0.0851
43	PCA	KNN	1.0000 ± 0.0000	0.7411 ± 0.0787
		MLP	0.8009 ± 0.1576	0.7214 ± 0.1497
		SVM	0.9314 ± 0.0129	0.7784 ± 0.0765
		RF	0.9644 ± 0.0098	0.8353 ± 0.0441
	TSVD	KNN	0.6971 ± 0.0258	0.6595 ± 0.0825
		MLP	0.9062 ± 0.0087	0.8119 ± 0.0684
		SVM	0.9015 ± 0.0115	0.7833 ± 0.0696
		RF	1.0000 ± 0.0000	0.8167 ± 0.0830
	KPCA	KNN	0.8019 ± 0.0178	0.6998 ± 0.0990
		MLP	0.9182 ± 0.0176	0.7976 ± 0.0647
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	0.9659 ± 0.0111	0.7742 ± 0.0704
	UMAP	KNN	0.7861 ± 0.0485	0.6991 ± 0.1272
		MLP	1.0000 ± 0.0000	0.7602 ± 0.0604
		SVM	0.9292 ± 0.0114	0.7883 ± 0.0476
		RF	0.9392 ± 0.0103	0.7788 ± 0.0611
	Ivis	KNN	0.9324 ± 0.0158	0.8403 ± 0.0655
		MLP	0.8155 ± 0.0213	0.7506 ± 0.1021
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	0.9130 ± 0.0136	0.7602 ± 0.1229
44	PCA	KNN	1.0000 ± 0.0000	0.7645 ± 0.0855
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.9314 ± 0.0119	0.7550 ± 0.0722
		RF	1.0000 ± 0.0000	0.8258 ± 0.0434
	TSVD	KNN	1.0000 ± 0.0000	0.6989 ± 0.0860
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.9345 ± 0.0102	0.7500 ± 0.0803
		RF	0.9172 ± 0.0160	0.7550 ± 0.1084
	KPCA	KNN	1.0000 ± 0.0000	0.7500 ± 0.0634
		MLP	0.9067 ± 0.0135	0.8165 ± 0.0631

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		SVM	0.9240 ± 0.0162	0.7790 ± 0.0836
		RF	0.9832 ± 0.0054	0.8260 ± 0.0568
	UMAP	KNN	0.7668 ± 0.0569	0.6894 ± 0.1141
		MLP	1.0000 ± 0.0000	0.6565 ± 0.1011
		SVM	1.0000 ± 0.0000	0.7323 ± 0.0897
		RF	0.9995 ± 0.0017	0.7881 ± 0.0540
	Ivis	KNN	0.9832 ± 0.0077	0.8214 ± 0.0809
		MLP	0.9953 ± 0.0046	0.8355 ± 0.0915
		SVM	0.9911 ± 0.0070	0.8262 ± 0.1081
		RF	1.0000 ± 0.0000	0.7883 ± 0.0791
45	PCA	KNN	0.7935 ± 0.0211	0.6937 ± 0.0653
		MLP	0.9188 ± 0.0133	0.7870 ± 0.0944
		SVM	0.9151 ± 0.0076	0.7831 ± 0.0839
		RF	1.0000 ± 0.0000	0.8258 ± 0.0434
	TSVD	KNN	0.7249 ± 0.0413	0.6552 ± 0.0796
		MLP	0.9093 ± 0.0048	0.8022 ± 0.0571
		SVM	0.9256 ± 0.0109	0.7550 ± 0.0811
		RF	1.0000 ± 0.0000	0.8019 ± 0.0889
	KPCA	KNN	0.7715 ± 0.0123	0.7316 ± 0.0755
		MLP	1.0000 ± 0.0000	0.7548 ± 0.0794
		SVM	0.9450 ± 0.0113	0.8121 ± 0.0828
		RF	1.0000 ± 0.0000	0.8310 ± 0.0848
	UMAP	KNN	0.9932 ± 0.0074	0.7561 ± 0.0879
		MLP	0.9292 ± 0.0103	0.7835 ± 0.0615
		SVM	1.0000 ± 0.0000	0.6610 ± 0.0965
		RF	1.0000 ± 0.0000	0.7457 ± 0.0481
	Ivis	KNN	0.7547 ± 0.0208	0.7084 ± 0.0873
		MLP	1.0000 ± 0.0000	0.8165 ± 0.0448
		SVM	0.9340 ± 0.0140	0.7929 ± 0.0799
		RF	1.0000 ± 0.0000	0.7885 ± 0.0843
46	PCA	KNN	0.6001 ± 0.0264	0.6089 ± 0.0457

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		MLP	0.8884 ± 0.0164	0.7690 ± 0.0673
		SVM	0.9266 ± 0.0125	0.7645 ± 0.0822
		RF	1.0000 ± 0.0000	0.8398 ± 0.0742
	TSVD	KNN	1.0000 ± 0.0000	0.7504 ± 0.0681
		MLP	0.9104 ± 0.0093	0.8115 ± 0.0738
		SVM	0.9319 ± 0.0104	0.7643 ± 0.0731
		RF	1.0000 ± 0.0000	0.7883 ± 0.1032
	KPCA	KNN	0.8333 ± 0.0139	0.7645 ± 0.0569
		MLP	0.5000 ± 0.0025	0.5000 ± 0.0224
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	0.9880 ± 0.0066	0.8262 ± 0.0751
	UMAP	KNN	1.0000 ± 0.0000	0.7835 ± 0.0651
		MLP	0.9853 ± 0.0033	0.7604 ± 0.0706
		SVM	1.0000 ± 0.0000	0.6610 ± 0.0965
		RF	0.9350 ± 0.0103	0.7835 ± 0.0615
	Ivis	KNN	1.0000 ± 0.0000	0.7312 ± 0.0838
		MLP	0.9712 ± 0.0289	0.8258 ± 0.0801
		SVM	0.9990 ± 0.0022	0.7976 ± 0.0727
		RF	1.0000 ± 0.0000	0.8026 ± 0.0920
47	PCA	KNN	1.0000 ± 0.0000	0.7028 ± 0.0673
		MLP	0.5021 ± 0.0011	0.4810 ± 0.0100
		SVM	0.5405 ± 0.0800	0.5128 ± 0.0835
		RF	0.9921 ± 0.0057	0.8494 ± 0.0690
	TSVD	KNN	0.6258 ± 0.0297	0.5652 ± 0.0705
		MLP	0.9010 ± 0.0056	0.7974 ± 0.0702
		SVM	0.9334 ± 0.0098	0.7597 ± 0.0707
		RF	1.0000 ± 0.0000	0.8258 ± 0.0769
	KPCA	KNN	1.0000 ± 0.0000	0.7792 ± 0.0657
		MLP	0.9408 ± 0.0133	0.7974 ± 0.0734
		SVM	0.9240 ± 0.0139	0.7268 ± 0.1145
		RF	1.0000 ± 0.0000	0.8160 ± 0.0850

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
	UMAP	KNN	1.0000 ± 0.0000	0.7312 ± 0.0588
		MLP	0.9859 ± 0.0025	0.6942 ± 0.0947
		SVM	0.9188 ± 0.0146	0.7838 ± 0.0640
		RF	1.0000 ± 0.0000	0.7831 ± 0.0634
	Ivis	KNN	1.0000 ± 0.0000	0.7554 ± 0.0962
		MLP	0.9879 ± 0.0096	0.8117 ± 0.0825
		SVM	0.9722 ± 0.0202	0.7788 ± 0.0685
		RF	0.9990 ± 0.0022	0.7693 ± 0.0694
48	PCA	KNN	1.0000 ± 0.0000	0.6792 ± 0.0435
		MLP	0.5252 ± 0.0069	0.4861 ± 0.0419
		SVM	0.9324 ± 0.0119	0.7600 ± 0.0661
		RF	1.0000 ± 0.0000	0.8355 ± 0.0765
	TSVD	KNN	1.0000 ± 0.0000	0.6894 ± 0.1321
		MLP	0.9298 ± 0.0093	0.8160 ± 0.0570
		SVM	0.9308 ± 0.0097	0.7595 ± 0.0748
		RF	0.9885 ± 0.0059	0.8121 ± 0.0895
	KPCA	KNN	1.0000 ± 0.0000	0.6699 ± 0.0537
		MLP	0.8569 ± 0.0141	0.7747 ± 0.0975
		SVM	0.9287 ± 0.0126	0.7643 ± 0.0731
		RF	1.0000 ± 0.0000	0.8400 ± 0.0664
	UMAP	KNN	0.9303 ± 0.0111	0.7931 ± 0.0610
		MLP	1.0000 ± 0.0000	0.6184 ± 0.0687
		SVM	0.9801 ± 0.0085	0.7690 ± 0.0642
		RF	1.0000 ± 0.0000	0.6565 ± 0.1011
	Ivis	KNN	1.0000 ± 0.0000	0.8123 ± 0.1054
		MLP	0.9817 ± 0.0114	0.8260 ± 0.0689
		SVM	0.9361 ± 0.0104	0.8074 ± 0.0880
		RF	0.9780 ± 0.0081	0.8117 ± 0.0796
49	PCA	KNN	1.0000 ± 0.0000	0.7749 ± 0.1200
		MLP	0.9083 ± 0.0066	0.7922 ± 0.0567
		SVM	1.0000 ± 0.0000	0.5474 ± 0.0420

Continued on next page

Table 2 – Mean accuracy of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Train score	Test score
Dimensionality	Projector	Classifier		
		RF	1.0000 ± 0.0000	0.8305 ± 0.0704
	TSVD	KNN	1.0000 ± 0.0000	0.6318 ± 0.0670
		MLP	0.9036 ± 0.0055	0.7974 ± 0.0539
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.8121 ± 0.0868
	KPCA	KNN	1.0000 ± 0.0000	0.7277 ± 0.1003
		MLP	0.9141 ± 0.0133	0.8069 ± 0.0673
		SVM	0.5021 ± 0.0011	0.4810 ± 0.0100
		RF	1.0000 ± 0.0000	0.8351 ± 0.0594
	UMAP	KNN	1.0000 ± 0.0000	0.6658 ± 0.0926
		MLP	0.0000 ± 0.0000	0.0000 ± 0.0000
		SVM	1.0000 ± 0.0000	0.7461 ± 0.0767
		RF	1.0000 ± 0.0000	0.7457 ± 0.0531
	Ivis	KNN	0.8805 ± 0.0219	0.7693 ± 0.0670
		MLP	1.0000 ± 0.0000	0.8210 ± 0.0573
		SVM	0.9392 ± 0.0102	0.8022 ± 0.0724
		RF	0.9287 ± 0.0255	0.8171 ± 0.0833

APPENDIX

B

CROSS-VALIDATION TIMES

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
50	—	KNN	0.0051 ± 0.0016	0.0100 ± 0.0036
		MLP	2.0270 ± 0.1369	0.0027 ± 0.0036
		SVM	0.0985 ± 0.0258	0.0013 ± 0.0002
		RF	0.2146 ± 0.0203	0.0356 ± 0.0184
2	PCA	KNN	0.0080 ± 0.0017	0.0020 ± 0.0002
		MLP	0.3705 ± 0.0289	0.0018 ± 0.0004
		SVM	0.1437 ± 0.0694	0.0018 ± 0.0013
		RF	0.2543 ± 0.0215	0.0228 ± 0.0013
	TSVD	KNN	0.0092 ± 0.0012	0.0022 ± 0.0002
		MLP	0.4677 ± 0.0278	0.0022 ± 0.0002
		SVM	0.1421 ± 0.0433	0.0019 ± 0.0017
		RF	0.6125 ± 0.0334	0.0490 ± 0.0063
	KPCA	KNN	0.0570 ± 0.0086	0.0279 ± 0.0039
		MLP	0.3445 ± 0.0376	0.0490 ± 0.0292
		SVM	0.1147 ± 0.0110	0.0425 ± 0.0123
		RF	0.2794 ± 0.0402	0.0659 ± 0.0175
	UMAP	KNN	2.4852 ± 0.2446	1.5056 ± 0.2197
		MLP	3.7267 ± 0.1962	1.8610 ± 0.1700

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	1.9406 ± 0.2242	1.4486 ± 0.2210
		RF	2.6243 ± 0.2281	1.3782 ± 0.2116
	Ivis	KNN	263.8681 ± 76.3864	0.7927 ± 0.2518
		MLP	32.2996 ± 7.6062	0.5360 ± 0.1722
		SVM	333.0231 ± 124.4573	0.8614 ± 0.2533
		RF	327.53 ± 94.6619	1.4752 ± 0.1677
3	PCA	KNN	0.0082 ± 0.0015	0.0032 ± 0.0007
		MLP	1.5443 ± 0.0973	0.0021 ± 0.0001
		SVM	0.0928 ± 0.0220	0.0013 ± 0.0002
		RF	0.2534 ± 0.0228	0.0315 ± 0.0059
	TSVD	KNN	0.0078 ± 0.0012	0.0022 ± 0.0003
		MLP	0.5159 ± 0.0649	0.0023 ± 0.0009
		SVM	0.0176 ± 0.0033	0.0015 ± 0.0003
		RF	0.4203 ± 0.0352	0.0527 ± 0.0193
	KPCA	KNN	0.0774 ± 0.0144	0.0364 ± 0.0114
		MLP	0.4898 ± 0.0896	0.0650 ± 0.0300
		SVM	0.0261 ± 0.0035	0.0328 ± 0.0145
		RF	0.4329 ± 0.0186	0.0539 ± 0.0036
	UMAP	KNN	3.5638 ± 0.2369	1.4701 ± 0.1589
		MLP	2.5560 ± 0.2265	2.0631 ± 0.1659
		SVM	5.0432 ± 0.5012	1.8808 ± 0.2254
		RF	4.7565 ± 0.1884	1.6826 ± 0.1002
	Ivis	KNN	255.3325 ± 49.0265	0.5076 ± 0.2948
		MLP	1.8062 ± 0.7161	0.3298 ± 0.3063
		SVM	1.8930 ± 0.3650	0.6887 ± 0.2262
		RF	91.7132 ± 15.638	0.3157 ± 0.1490
4	PCA	KNN	0.0092 ± 0.0029	0.0030 ± 0.0005
		MLP	0.4324 ± 0.0344	0.0018 ± 0.0002
		SVM	0.0160 ± 0.0025	0.0011 ± 0.0003
		RF	0.4941 ± 0.0320	0.0493 ± 0.0105

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0070 ± 0.0016	0.0033 ± 0.0001
		MLP	0.8361 ± 0.2144	0.0018 ± 0.0002
		SVM	0.0817 ± 0.0168	0.0011 ± 0.0001
		RF	0.2223 ± 0.0196	0.0315 ± 0.0073
	KPCA	KNN	0.0561 ± 0.0089	0.0251 ± 0.0076
		MLP	1.5071 ± 0.0938	0.0483 ± 0.0274
		SVM	0.0186 ± 0.0049	0.0317 ± 0.0166
		RF	0.6836 ± 0.0418	0.1011 ± 0.0237
	UMAP	KNN	3.9770 ± 0.2612	1.6753 ± 0.2142
		MLP	4.5270 ± 0.2803	1.6970 ± 0.2068
		SVM	4.4077 ± 0.2523	2.1146 ± 0.2193
		RF	2.2566 ± 0.1342	1.7050 ± 0.1770
	Ivis	KNN	191.7105 ± 46.037	0.8930 ± 0.1680
		MLP	15.0756 ± 4.1370	0.2081 ± 0.0627
		SVM	329.1117 ± 53.8819	0.5415 ± 0.1599
		RF	20.0378 ± 2.7747	0.1644 ± 0.0297
5	PCA	KNN	0.0072 ± 0.0014	0.0031 ± 0.0005
		MLP	1.0256 ± 0.0523	0.0018 ± 0.0002
		SVM	0.0482 ± 0.0060	0.0015 ± 0.0001
		RF	0.6133 ± 0.0387	0.0685 ± 0.0360
	TSVD	KNN	0.0062 ± 0.0008	0.0031 ± 0.0006
		MLP	0.5221 ± 0.0213	0.0018 ± 0.0001
		SVM	0.0309 ± 0.0028	0.0013 ± 0.0002
		RF	0.4897 ± 0.0327	0.0586 ± 0.0230
	KPCA	KNN	0.0552 ± 0.0065	0.0259 ± 0.0056
		MLP	0.9126 ± 0.0330	0.0683 ± 0.0317
		SVM	0.0281 ± 0.0050	0.0280 ± 0.0066
		RF	0.6737 ± 0.0434	0.0768 ± 0.0138
	UMAP	KNN	0.0123 ± 0.0008	1.6466 ± 0.1277
		MLP	3.9355 ± 0.2736	1.6736 ± 0.1397

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	1.7567 ± 0.1784	1.6342 ± 0.1731
		RF	3.0880 ± 0.2434	1.6439 ± 0.1617
	Ivis	KNN	33.5962 ± 7.5007	0.8368 ± 0.3499
		MLP	275.5294 ± 64.9767	0.7377 ± 0.1718
		SVM	332.4185 ± 73.5912	0.6152 ± 0.1889
		RF	120.9599 ± 47.8915	1.0816 ± 0.3638
6	PCA	KNN	0.0096 ± 0.0021	0.0021 ± 0.0004
		MLP	1.6681 ± 0.1619	0.0019 ± 0.0002
		SVM	0.0346 ± 0.0149	0.0016 ± 0.0003
		RF	0.5932 ± 0.0418	0.0886 ± 0.0746
	TSVD	KNN	0.0074 ± 0.0013	0.0033 ± 0.0008
		MLP	0.8021 ± 0.0364	0.0026 ± 0.0006
		SVM	0.0272 ± 0.0026	0.0016 ± 0.0001
		RF	0.5364 ± 0.0305	0.0567 ± 0.0163
	KPCA	KNN	0.0633 ± 0.0099	0.0355 ± 0.0149
		MLP	2.3179 ± 0.1103	0.0539 ± 0.0162
		SVM	0.0244 ± 0.0026	0.0314 ± 0.0049
		RF	0.2724 ± 0.0449	0.0518 ± 0.0128
	UMAP	KNN	5.2458 ± 0.3221	1.6722 ± 0.1384
		MLP	3.4771 ± 0.2392	1.6945 ± 0.2097
		SVM	3.5193 ± 0.2648	1.7648 ± 0.2279
		RF	3.4583 ± 0.1623	1.6788 ± 0.1613
	Ivis	KNN	285.7485 ± 55.488	0.7092 ± 0.1911
		MLP	100.2301 ± 17.016	0.2682 ± 0.0716
		SVM	324.0569 ± 45.9722	0.5792 ± 0.2136
		RF	43.7625 ± 12.1647	1.1318 ± 0.2334
7	PCA	KNN	0.0082 ± 0.0013	0.0025 ± 0.0004
		MLP	0.6518 ± 0.1061	0.0041 ± 0.0039
		SVM	0.0262 ± 0.0036	0.0014 ± 0.0003
		RF	0.6236 ± 0.0327	0.0577 ± 0.0134

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0085 ± 0.0018	0.0040 ± 0.0010
		MLP	0.9671 ± 0.0736	0.0014 ± 0.0006
		SVM	0.0196 ± 0.0007	0.0017 ± 0.0006
		RF	0.2664 ± 0.0172	0.0369 ± 0.0199
	KPCA	KNN	0.0561 ± 0.0120	0.0345 ± 0.0124
		MLP	1.3541 ± 0.0709	0.0830 ± 0.0513
		SVM	0.0666 ± 0.0207	0.0338 ± 0.0142
		RF	0.6385 ± 0.0403	0.0706 ± 0.0125
	UMAP	KNN	1.8529 ± 0.2110	1.6138 ± 0.1395
		MLP	4.2035 ± 0.3319	1.8766 ± 0.2171
		SVM	2.7500 ± 0.2685	1.6162 ± 0.2219
		RF	5.2229 ± 0.2500	2.3172 ± 0.0865
	Ivis	KNN	88.3916 ± 17.4931	0.2128 ± 0.0732
		MLP	270.7918 ± 70.0257	0.7980 ± 0.1475
		SVM	91.7366 ± 21.3337	0.7619 ± 0.1239
		RF	59.3767 ± 12.0753	0.6333 ± 0.6233
8	PCA	KNN	0.0112 ± 0.0035	0.0029 ± 0.0002
		MLP	0.5486 ± 0.0416	0.0020 ± 0.0006
		SVM	0.0363 ± 0.0054	0.0012 ± 0.0004
		RF	0.3219 ± 0.0221	0.0351 ± 0.0094
	TSVD	KNN	0.0087 ± 0.0024	0.0024 ± 0.0004
		MLP	1.2516 ± 0.0828	0.0020 ± 0.0001
		SVM	0.0179 ± 0.0019	0.0018 ± 0.0019
		RF	0.3777 ± 0.0314	0.0400 ± 0.0102
	KPCA	KNN	0.0663 ± 0.0171	0.0287 ± 0.0082
		MLP	5.2736 ± 0.3505	0.0768 ± 0.0379
		SVM	0.0729 ± 0.0099	0.0341 ± 0.0132
		RF	0.5321 ± 0.0387	0.0803 ± 0.0170
	UMAP	KNN	1.7704 ± 0.2989	1.5567 ± 0.1557
		MLP	2.4454 ± 0.1665	2.1592 ± 0.2726

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	6.8695 ± 0.4281	2.2992 ± 0.3297
		RF	1.7991 ± 0.0915	1.6834 ± 0.3034
	Ivis	KNN	175.5464 ± 58.427	0.8276 ± 0.3260
		MLP	290.3302 ± 68.2255	0.8213 ± 0.2817
		SVM	290.3512 ± 79.953	0.7171 ± 0.0896
		RF	23.5325 ± 5.8888	1.0873 ± 0.3319
9	PCA	KNN	0.0083 ± 0.0009	0.0032 ± 0.0003
		MLP	2.2607 ± 0.1006	0.0021 ± 0.0002
		SVM	0.0237 ± 0.0040	0.0012 ± 0.0003
		RF	0.3382 ± 0.0356	0.0569 ± 0.0256
	TSVD	KNN	0.0100 ± 0.0028	0.0022 ± 0.0004
		MLP	1.0632 ± 0.0351	0.0020 ± 0.0001
		SVM	0.0373 ± 0.0065	0.0015 ± 0.0000
		RF	0.4198 ± 0.0364	0.0372 ± 0.0084
	KPCA	KNN	0.0624 ± 0.0118	0.0335 ± 0.0123
		MLP	6.4930 ± 0.4486	0.0487 ± 0.0217
		SVM	0.1683 ± 0.0181	0.0290 ± 0.0142
		RF	0.6067 ± 0.0299	0.0767 ± 0.0134
	UMAP	KNN	5.2737 ± 0.2249	1.9617 ± 0.1648
		MLP	0.4606 ± 0.1004	2.6232 ± 0.1888
		SVM	2.8709 ± 0.2670	1.5770 ± 0.1674
		RF	6.2361 ± 0.2075	2.1515 ± 0.1574
	Ivis	KNN	280.3055 ± 55.5147	0.6485 ± 0.1519
		MLP	121.5438 ± 71.8424	0.5954 ± 0.2477
		SVM	189.7145 ± 67.3603	0.8215 ± 0.3220
		RF	68.2415 ± 18.9854	1.1669 ± 0.2012
10	PCA	KNN	0.0116 ± 0.0013	0.0036 ± 0.0002
		MLP	0.7172 ± 0.0826	0.0022 ± 0.0003
		SVM	0.1994 ± 0.0315	0.0013 ± 0.0001
		RF	0.3531 ± 0.0264	0.0304 ± 0.0066

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0485 ± 0.0209	0.0021 ± 0.0001
		MLP	2.1977 ± 0.1509	0.0017 ± 0.0002
		SVM	0.0210 ± 0.0028	0.0012 ± 0.0003
		RF	0.2177 ± 0.0181	0.0258 ± 0.0079
	KPCA	KNN	0.0084 ± 0.0010	0.0407 ± 0.0096
		MLP	3.2896 ± 0.1627	0.0851 ± 0.0359
		SVM	0.2776 ± 0.0372	0.0333 ± 0.0134
		RF	0.2710 ± 0.0261	0.0655 ± 0.0135
	UMAP	KNN	2.6719 ± 0.2339	2.2999 ± 0.1989
		MLP	2.6051 ± 0.2719	2.6354 ± 0.2492
		SVM	2.4320 ± 0.1536	1.8330 ± 0.1580
		RF	2.1377 ± 0.1780	1.5984 ± 0.2119
	Ivis	KNN	101.6583 ± 23.7489	0.7020 ± 0.1724
		MLP	1.6331 ± 1.3779	0.2250 ± 0.0799
		SVM	0.6946 ± 0.0663	0.2158 ± 0.0244
		RF	90.771 ± 24.5667	0.3432 ± 0.1696
11	PCA	KNN	0.0088 ± 0.0008	0.0032 ± 0.0003
		MLP	2.0487 ± 0.4248	0.0022 ± 0.0002
		SVM	0.2591 ± 0.0342	0.0013 ± 0.0001
		RF	0.5951 ± 0.0435	0.0511 ± 0.0134
	TSVD	KNN	0.0080 ± 0.0027	0.0027 ± 0.0011
		MLP	6.1298 ± 1.0033	0.0020 ± 0.0001
		SVM	0.0689 ± 0.0116	0.0019 ± 0.0013
		RF	0.5804 ± 0.0383	0.0469 ± 0.0065
	KPCA	KNN	0.0744 ± 0.0103	0.0379 ± 0.0095
		MLP	1.8584 ± 0.0986	0.0202 ± 0.0017
		SVM	0.2262 ± 0.0345	0.0238 ± 0.0094
		RF	0.6140 ± 0.0637	0.0697 ± 0.0105
	UMAP	KNN	1.7290 ± 0.1581	1.5741 ± 0.1742
		MLP	3.6342 ± 0.3217	2.3606 ± 0.1880

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	2.2306 ± 0.1751	2.0415 ± 0.2157
		RF	5.7349 ± 0.2292	1.8281 ± 0.1208
	Ivis	KNN	94.5166 ± 21.6701	0.2858 ± 0.1275
		MLP	25.973 ± 7.0730	0.2729 ± 0.1336
		SVM	127.3769 ± 32.8844	0.8198 ± 0.2674
		RF	153.0624 ± 54.7699	0.9599 ± 0.1264
12	PCA	KNN	0.0095 ± 0.0005	0.0039 ± 0.0004
		MLP	3.3643 ± 0.1424	0.0020 ± 0.0001
		SVM	0.2621 ± 0.0339	0.0013 ± 0.0002
		RF	0.4121 ± 0.0533	0.0307 ± 0.0032
	TSVD	KNN	0.0075 ± 0.0012	0.0020 ± 0.0003
		MLP	1.6487 ± 0.0765	0.0031 ± 0.0038
		SVM	0.0199 ± 0.0013	0.0014 ± 0.0002
		RF	0.6172 ± 0.0456	0.0516 ± 0.0078
	KPCA	KNN	0.0107 ± 0.0029	0.0271 ± 0.0088
		MLP	1.4372 ± 0.0785	0.0455 ± 0.0168
		SVM	0.0842 ± 0.0118	0.0339 ± 0.0112
		RF	0.6078 ± 0.0744	0.0743 ± 0.0183
	UMAP	KNN	1.7506 ± 0.1992	1.5181 ± 0.1786
		MLP	5.5387 ± 0.4641	2.0342 ± 0.2454
		SVM	1.8369 ± 0.1385	1.5491 ± 0.2241
		RF	3.1919 ± 0.1572	1.6498 ± 0.1366
	Ivis	KNN	2.3928 ± 5.8922	0.2372 ± 0.0306
		MLP	189.9352 ± 35.885	0.2680 ± 0.1064
		SVM	34.028 ± 6.4241	0.6623 ± 0.1894
		RF	164.1049 ± 17.5547	1.1591 ± 0.2461
13	PCA	KNN	0.0109 ± 0.0065	0.0051 ± 0.0015
		MLP	2.6919 ± 0.0896	0.0018 ± 0.0003
		SVM	0.0375 ± 0.0152	0.0018 ± 0.0003
		RF	0.6071 ± 0.0491	0.0476 ± 0.0080

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0077 ± 0.0015	0.0046 ± 0.0005
		MLP	2.9092 ± 0.2693	0.0019 ± 0.0001
		SVM	0.0281 ± 0.0045	0.0018 ± 0.0003
		RF	0.3101 ± 0.0285	0.0245 ± 0.0059
	KPCA	KNN	0.0655 ± 0.0127	0.0308 ± 0.0086
		MLP	1.7977 ± 0.0722	0.0915 ± 0.0283
		SVM	0.1594 ± 0.0235	0.0268 ± 0.0083
		RF	0.2710 ± 0.0171	0.0678 ± 0.0094
	UMAP	KNN	5.1389 ± 0.4230	1.7913 ± 0.2456
		MLP	5.0277 ± 0.3880	1.7263 ± 0.2230
		SVM	1.7891 ± 0.1169	1.5917 ± 0.1475
		RF	2.2708 ± 0.2065	2.1479 ± 0.1266
	Ivis	KNN	174.2943 ± 45.8891	0.7163 ± 0.2058
		MLP	306.3958 ± 71.2342	0.8153 ± 0.2269
		SVM	255.7838 ± 37.8483	0.7247 ± 0.2452
		RF	393.5508 ± 73.8103	1.4723 ± 0.2682
14	PCA	KNN	0.0098 ± 0.0020	0.0025 ± 0.0006
		MLP	0.6388 ± 0.0639	0.0020 ± 0.0002
		SVM	0.0256 ± 0.0061	0.0015 ± 0.0001
		RF	0.2237 ± 0.0209	0.0317 ± 0.0093
	TSVD	KNN	0.0066 ± 0.0008	0.0062 ± 0.0023
		MLP	0.5302 ± 0.0680	0.0021 ± 0.0011
		SVM	0.0288 ± 0.0053	0.0018 ± 0.0002
		RF	0.4333 ± 0.0309	0.0524 ± 0.0229
	KPCA	KNN	0.0080 ± 0.0012	0.0332 ± 0.0112
		MLP	0.6818 ± 0.0249	0.0600 ± 0.0358
		SVM	0.0798 ± 0.0077	0.0327 ± 0.0194
		RF	0.4698 ± 0.0262	0.0756 ± 0.0207
	UMAP	KNN	2.7156 ± 0.2054	1.5410 ± 0.1327
		MLP	3.9189 ± 0.1901	2.8760 ± 0.2718

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	1.9781 ± 0.1730	1.6596 ± 0.1708
		RF	2.9278 ± 0.1772	0.0000 ± 0.0000
	Ivis	KNN	3.7443 ± 9.8181	0.2062 ± 0.0521
		MLP	152.431 ± 38.0058	0.5659 ± 0.2461
		SVM	26.2192 ± 71.9577	0.9022 ± 0.4323
		RF	255.1362 ± 43.1554	0.4066 ± 0.2604
15	PCA	KNN	0.0095 ± 0.0012	0.0047 ± 0.0029
		MLP	4.9567 ± 0.5174	0.0020 ± 0.0002
		SVM	0.0400 ± 0.0113	0.0017 ± 0.0001
		RF	0.2769 ± 0.0271	0.0335 ± 0.0091
	TSVD	KNN	0.0074 ± 0.0010	0.0041 ± 0.0006
		MLP	1.8766 ± 0.1987	0.0021 ± 0.0002
		SVM	0.0202 ± 0.0004	0.0014 ± 0.0001
		RF	0.3151 ± 0.0229	0.0296 ± 0.0054
	KPCA	KNN	0.0712 ± 0.0193	0.0299 ± 0.0133
		MLP	0.5047 ± 0.0364	0.0489 ± 0.0262
		SVM	0.3578 ± 0.0459	0.0411 ± 0.0196
		RF	0.6535 ± 0.0437	0.0702 ± 0.0092
	UMAP	KNN	1.6847 ± 0.1809	1.6019 ± 0.1438
		MLP	3.7954 ± 0.2377	1.9546 ± 0.3129
		SVM	2.2166 ± 0.1689	1.4348 ± 0.2356
		RF	2.1054 ± 0.1412	1.8250 ± 0.1056
	Ivis	KNN	21.3008 ± 3.9834	0.3586 ± 0.1402
		MLP	2.7510 ± 1.2476	0.3298 ± 0.3217
		SVM	83.9684 ± 15.3701	0.4274 ± 0.1260
		RF	282.7674 ± 63.2912	1.3825 ± 0.1864
16	PCA	KNN	0.0091 ± 0.0015	0.0020 ± 0.0002
		MLP	0.8587 ± 0.0598	0.0034 ± 0.0029
		SVM	0.0244 ± 0.0065	0.0023 ± 0.0025
		RF	0.3326 ± 0.0338	0.0401 ± 0.0126

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0080 ± 0.0030	0.0066 ± 0.0022
		MLP	1.9692 ± 0.0891	0.0010 ± 0.0000
		SVM	0.0348 ± 0.0116	0.0014 ± 0.0002
		RF	0.2912 ± 0.0283	0.0344 ± 0.0085
	KPCA	KNN	0.0754 ± 0.0138	0.0438 ± 0.0134
		MLP	0.3556 ± 0.0282	0.0529 ± 0.0198
		SVM	0.0252 ± 0.0022	0.0301 ± 0.0088
		RF	0.2615 ± 0.0332	0.0538 ± 0.0126
	UMAP	KNN	1.8249 ± 0.2881	1.4812 ± 0.1936
		MLP	2.4250 ± 0.2575	1.6323 ± 0.1498
		SVM	2.8275 ± 0.2489	1.5350 ± 0.1408
		RF	5.1069 ± 0.2343	2.0546 ± 0.3782
	Ivis	KNN	32.0349 ± 5.5832	0.3642 ± 0.1219
		MLP	114.2056 ± 34.7846	0.5179 ± 0.0915
		SVM	53.2398 ± 14.4421	0.2177 ± 0.0275
		RF	400.3298 ± 102.7405	1.5827 ± 0.6043
17	PCA	KNN	0.0090 ± 0.0015	0.0048 ± 0.0002
		MLP	3.7609 ± 0.3246	0.0019 ± 0.0000
		SVM	0.0328 ± 0.0071	0.0014 ± 0.0003
		RF	0.6260 ± 0.0406	0.0507 ± 0.0087
	TSVD	KNN	0.0061 ± 0.0007	0.0018 ± 0.0002
		MLP	3.7858 ± 0.4089	0.0019 ± 0.0001
		SVM	0.0219 ± 0.0009	0.0015 ± 0.0001
		RF	0.6323 ± 0.0337	0.0454 ± 0.0113
	KPCA	KNN	0.0742 ± 0.0131	0.0383 ± 0.0108
		MLP	1.8617 ± 0.1000	0.0879 ± 0.0496
		SVM	0.2730 ± 0.0277	0.0346 ± 0.0126
		RF	0.2705 ± 0.0332	0.0538 ± 0.0155
	UMAP	KNN	1.8856 ± 0.2158	3.0141 ± 0.2239
		MLP	2.5186 ± 0.2183	1.8339 ± 0.2506

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	6.9339 ± 0.3917	2.1755 ± 0.2636
		RF	3.5980 ± 0.1691	1.6075 ± 0.0975
	Ivis	KNN	348.4073 ± 53.9086	0.7400 ± 0.2844
		MLP	283.3511 ± 76.1751	0.8617 ± 0.1475
		SVM	98.5945 ± 29.6863	0.5135 ± 0.1169
		RF	187.8674 ± 34.8004	1.1487 ± 0.2559
18	PCA	KNN	0.0097 ± 0.0042	0.0057 ± 0.0007
		MLP	2.4976 ± 0.0827	0.0020 ± 0.0004
		SVM	0.2080 ± 0.0552	0.0024 ± 0.0032
		RF	0.3914 ± 0.0374	0.0466 ± 0.0117
	TSVD	KNN	0.0070 ± 0.0006	0.0058 ± 0.0004
		MLP	0.4224 ± 0.0397	0.0018 ± 0.0002
		SVM	0.0247 ± 0.0039	0.0013 ± 0.0002
		RF	0.6003 ± 0.0187	0.0511 ± 0.0079
	KPCA	KNN	0.0688 ± 0.0065	0.0361 ± 0.0095
		MLP	2.8708 ± 0.0641	0.0679 ± 0.0297
		SVM	0.0949 ± 0.0218	0.0313 ± 0.0120
		RF	0.6369 ± 0.0486	0.0899 ± 0.0189
	UMAP	KNN	4.5187 ± 0.1992	2.1175 ± 0.2714
		MLP	3.2002 ± 0.3230	2.0048 ± 0.2429
		SVM	2.1417 ± 0.1511	1.6199 ± 0.1932
		RF	3.3129 ± 0.2814	1.7085 ± 0.1754
	Ivis	KNN	0.4185 ± 0.0385	0.3711 ± 0.1291
		MLP	309.1511 ± 111.796	0.8533 ± 0.2068
		SVM	0.3594 ± 0.0367	0.1746 ± 0.0201
		RF	115.464 ± 31.043	1.0325 ± 0.3977
19	PCA	KNN	0.0099 ± 0.0021	0.0061 ± 0.0003
		MLP	0.7825 ± 0.0601	0.0023 ± 0.0003
		SVM	0.0345 ± 0.0045	0.0017 ± 0.0002
		RF	0.2104 ± 0.0289	0.0655 ± 0.1005

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0079 ± 0.0030	0.0050 ± 0.0005
		MLP	3.7421 ± 0.4309	0.0018 ± 0.0001
		SVM	0.3326 ± 0.0433	0.0011 ± 0.0002
		RF	0.3449 ± 0.0402	0.0453 ± 0.0185
	KPCA	KNN	0.0094 ± 0.0019	0.0291 ± 0.0099
		MLP	0.8569 ± 0.3691	0.0711 ± 0.0381
		SVM	0.0853 ± 0.0077	0.0333 ± 0.0118
		RF	0.2845 ± 0.0342	0.0551 ± 0.0059
	UMAP	KNN	1.8335 ± 0.2132	1.6877 ± 0.2341
		MLP	2.6043 ± 0.1737	1.5655 ± 0.2311
		SVM	1.9718 ± 0.2492	3.1941 ± 0.4134
		RF	9.3838 ± 0.1675	3.0754 ± 0.1512
	Ivis	KNN	49.6436 ± 14.495	0.6786 ± 0.1692
		MLP	7.6324 ± 1.3693	0.8288 ± 0.3659
		SVM	127.7771 ± 62.5866	0.4887 ± 0.1673
		RF	384.1555 ± 114.8796	1.3470 ± 0.4130
20	PCA	KNN	0.0081 ± 0.0009	0.0052 ± 0.0004
		MLP	0.8500 ± 0.0764	0.0022 ± 0.0002
		SVM	0.0384 ± 0.0039	0.0017 ± 0.0001
		RF	0.4774 ± 0.0376	0.0314 ± 0.0054
	TSVD	KNN	0.0069 ± 0.0018	0.0059 ± 0.0005
		MLP	3.2466 ± 0.4639	0.0019 ± 0.0001
		SVM	0.0329 ± 0.0024	0.0016 ± 0.0002
		RF	0.4834 ± 0.0426	0.0502 ± 0.0149
	KPCA	KNN	0.0599 ± 0.0149	0.0302 ± 0.0088
		MLP	3.1023 ± 0.0963	0.0646 ± 0.0218
		SVM	0.3760 ± 0.0317	0.0470 ± 0.0219
		RF	0.5072 ± 0.0440	0.0710 ± 0.0112
	UMAP	KNN	1.9767 ± 0.1435	1.7964 ± 0.2043
		MLP	7.7786 ± 0.2628	2.5075 ± 0.1560

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	1.9509 ± 0.1403	2.0878 ± 0.1964
		RF	2.7312 ± 0.1959	1.6683 ± 0.0938
	Ivis	KNN	15.755 ± 3.9541	0.2102 ± 0.0357
		MLP	144.0386 ± 26.1517	0.2761 ± 0.1260
		SVM	37.3571 ± 8.0556	0.8582 ± 0.2813
		RF	103.6871 ± 17.7481	1.1119 ± 0.2396
21	PCA	KNN	0.0081 ± 0.0007	0.0085 ± 0.0063
		MLP	1.4112 ± 0.1324	0.0021 ± 0.0002
		SVM	0.0342 ± 0.0051	0.0016 ± 0.0002
		RF	0.4051 ± 0.0311	0.0463 ± 0.0095
	TSVD	KNN	0.0075 ± 0.0025	0.0017 ± 0.0002
		MLP	0.7973 ± 0.0995	0.0020 ± 0.0007
		SVM	0.0284 ± 0.0018	0.0016 ± 0.0001
		RF	0.2636 ± 0.0551	0.0344 ± 0.0098
	KPCA	KNN	0.0759 ± 0.0121	0.0378 ± 0.0084
		MLP	0.5977 ± 0.0473	0.0439 ± 0.0165
		SVM	0.0887 ± 0.0202	0.0335 ± 0.0166
		RF	0.5634 ± 0.0430	0.0594 ± 0.0070
	UMAP	KNN	9.5905 ± 0.3494	3.8658 ± 0.1922
		MLP	5.7731 ± 0.4122	2.0994 ± 0.1960
		SVM	3.3220 ± 0.2830	1.6491 ± 0.2008
		RF	2.6288 ± 0.1561	1.7170 ± 0.1832
	Ivis	KNN	238.4851 ± 34.779	0.5896 ± 0.2698
		MLP	270.6783 ± 49.7738	0.9142 ± 0.2805
		SVM	33.7209 ± 8.9934	0.4568 ± 0.2326
		RF	143.8766 ± 27.4806	1.0132 ± 0.2063
22	PCA	KNN	0.0099 ± 0.0014	0.0060 ± 0.0005
		MLP	3.0823 ± 0.2482	0.0016 ± 0.0003
		SVM	0.0333 ± 0.0017	0.0017 ± 0.0002
		RF	0.4675 ± 0.0242	0.0328 ± 0.0043

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0065 ± 0.0004	0.0065 ± 0.0002
		MLP	0.4971 ± 0.0201	0.0020 ± 0.0003
		SVM	0.3192 ± 0.0416	0.0015 ± 0.0006
		RF	0.3815 ± 0.0224	0.0406 ± 0.0155
	KPCA	KNN	0.0631 ± 0.0098	0.0327 ± 0.0106
		MLP	1.2372 ± 0.1215	0.0491 ± 0.0305
		SVM	0.0840 ± 0.0097	0.0467 ± 0.0143
		RF	0.2695 ± 0.0188	0.0602 ± 0.0179
	UMAP	KNN	3.1168 ± 0.1936	1.5985 ± 0.2233
		MLP	2.9869 ± 0.3776	1.8419 ± 0.2026
		SVM	10.6708 ± 0.4159	3.8696 ± 0.3092
		RF	9.0777 ± 0.2419	3.1340 ± 0.1866
	Ivis	KNN	147.4315 ± 43.3901	0.5284 ± 0.2010
		MLP	220.7913 ± 61.2446	0.7648 ± 0.3041
		SVM	0.7358 ± 0.0672	0.2136 ± 0.0441
		RF	193.5886 ± 68.9495	1.0621 ± 0.1778
23	PCA	KNN	0.0086 ± 0.0015	0.0066 ± 0.0008
		MLP	0.7274 ± 0.0733	0.0020 ± 0.0002
		SVM	0.2833 ± 0.0826	0.0013 ± 0.0003
		RF	0.5804 ± 0.0439	0.0485 ± 0.0096
	TSVD	KNN	0.0077 ± 0.0016	0.0021 ± 0.0003
		MLP	3.7717 ± 0.6762	0.0019 ± 0.0001
		SVM	0.1763 ± 0.0405	0.0019 ± 0.0017
		RF	0.4549 ± 0.0570	0.0402 ± 0.0093
	KPCA	KNN	0.0701 ± 0.0097	0.0350 ± 0.0098
		MLP	0.4201 ± 0.1189	0.0658 ± 0.0200
		SVM	0.0929 ± 0.0224	0.0397 ± 0.0229
		RF	0.6771 ± 0.0334	0.0934 ± 0.0318
	UMAP	KNN	3.8295 ± 0.1904	3.5888 ± 0.3221
		MLP	7.2011 ± 0.6214	4.0879 ± 0.1955

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	2.5363 ± 0.2703	1.7809 ± 0.2337
		RF	3.7006 ± 0.0955	1.7822 ± 0.2422
	Ivis	KNN	96.118 ± 144.3785	0.7451 ± 0.5191
		MLP	191.3403 ± 38.1834	0.7424 ± 0.2868
		SVM	419.3141 ± 75.0439	0.9373 ± 0.2611
		RF	49.5813 ± 9.8009	1.4942 ± 0.7420
24	PCA	KNN	0.0083 ± 0.0012	0.0071 ± 0.0004
		MLP	0.3964 ± 0.0259	0.0018 ± 0.0002
		SVM	0.0387 ± 0.0123	0.0024 ± 0.0019
		RF	0.2116 ± 0.0157	0.0204 ± 0.0024
	TSVD	KNN	0.0067 ± 0.0008	0.0067 ± 0.0013
		MLP	3.2374 ± 0.2394	0.0020 ± 0.0002
		SVM	0.0298 ± 0.0052	0.0014 ± 0.0002
		RF	0.4955 ± 0.0469	0.0502 ± 0.0073
	KPCA	KNN	0.0755 ± 0.0168	0.0409 ± 0.0131
		MLP	2.0017 ± 0.1170	0.0969 ± 0.0407
		SVM	0.0342 ± 0.0055	0.0277 ± 0.0119
		RF	0.4715 ± 0.0473	0.0795 ± 0.0232
	UMAP	KNN	1.8439 ± 0.1938	1.6107 ± 0.2333
		MLP	4.2703 ± 0.3365	1.1323 ± 0.2593
		SVM	6.3086 ± 0.1982	1.8423 ± 0.2254
		RF	4.1561 ± 0.3069	2.4953 ± 0.3526
	Ivis	KNN	1.9725 ± 0.3069	0.8848 ± 0.3202
		MLP	4.7931 ± 0.5538	1.0346 ± 0.4550
		SVM	135.213 ± 15.7159	0.2659 ± 0.1170
		RF	149.827 ± 39.3499	1.2081 ± 0.2946
25	PCA	KNN	0.0090 ± 0.0014	0.0071 ± 0.0018
		MLP	2.2447 ± 0.1334	0.0035 ± 0.0037
		SVM	0.0245 ± 0.0049	0.0014 ± 0.0003
		RF	0.2612 ± 0.0324	0.0305 ± 0.0100

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0083 ± 0.0020	0.0075 ± 0.0036
		MLP	0.9568 ± 0.0765	0.0042 ± 0.0071
		SVM	0.3765 ± 0.0798	0.0011 ± 0.0002
		RF	0.4302 ± 0.0322	0.0361 ± 0.0082
	KPCA	KNN	0.0728 ± 0.0177	0.0525 ± 0.0132
		MLP	1.2654 ± 0.0597	0.0514 ± 0.0158
		SVM	0.3557 ± 0.0744	0.0371 ± 0.0103
		RF	0.4570 ± 0.0478	0.0745 ± 0.0222
	UMAP	KNN	2.3138 ± 0.1798	1.5947 ± 0.2884
		MLP	3.8581 ± 0.2373	2.0420 ± 0.3638
		SVM	2.0531 ± 0.2588	2.1280 ± 0.1607
		RF	4.0282 ± 0.3038	1.7186 ± 0.2259
	Ivis	KNN	83.0747 ± 17.1162	0.4039 ± 0.0811
		MLP	323.2538 ± 88.9825	0.6646 ± 0.2138
		SVM	18.962 ± 11.0512	0.4245 ± 0.1147
		RF	18.2298 ± 4.0635	0.5493 ± 0.1656
26	PCA	KNN	0.0089 ± 0.0017	0.0085 ± 0.0020
		MLP	4.2336 ± 2.4901	0.0021 ± 0.0001
		SVM	0.0287 ± 0.0018	0.0017 ± 0.0001
		RF	0.5900 ± 0.0688	0.0485 ± 0.0090
	TSVD	KNN	0.0079 ± 0.0012	0.0064 ± 0.0005
		MLP	0.9642 ± 0.0478	0.0026 ± 0.0011
		SVM	0.0360 ± 0.0034	0.0018 ± 0.0002
		RF	0.6171 ± 0.0338	0.0536 ± 0.0076
	KPCA	KNN	0.0761 ± 0.0148	0.0459 ± 0.0094
		MLP	2.0859 ± 0.0493	0.0715 ± 0.0169
		SVM	0.4413 ± 0.0763	0.0291 ± 0.0138
		RF	0.2866 ± 0.0240	0.0517 ± 0.0076
	UMAP	KNN	2.2998 ± 0.2260	1.4807 ± 0.1629
		MLP	2.4283 ± 0.3556	1.7788 ± 0.2645

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	1.7040 ± 0.2009	0.0000 ± 0.0000
		RF	10.6968 ± 0.2423	4.0182 ± 0.4639
	Ivis	KNN	141.9707 ± 42.9443	0.7472 ± 0.1746
		MLP	2.2516 ± 0.1908	0.5644 ± 0.0737
		SVM	0.7742 ± 0.1021	0.2484 ± 0.0571
		RF	143.5333 ± 52.7387	1.2899 ± 0.4150
27	PCA	KNN	0.0094 ± 0.0011	0.0024 ± 0.0008
		MLP	0.5816 ± 0.0497	0.0020 ± 0.0002
		SVM	0.0381 ± 0.0090	0.0017 ± 0.0003
		RF	0.5577 ± 0.0424	0.0524 ± 0.0129
	TSVD	KNN	0.0067 ± 0.0007	0.0025 ± 0.0006
		MLP	4.9747 ± 0.6820	0.0024 ± 0.0014
		SVM	0.2018 ± 0.0480	0.0012 ± 0.0003
		RF	0.2992 ± 0.0239	0.0346 ± 0.0083
	KPCA	KNN	0.0660 ± 0.0072	0.0322 ± 0.0102
		MLP	0.4595 ± 0.0222	0.0649 ± 0.0323
		SVM	0.2586 ± 0.0666	0.0452 ± 0.0178
		RF	0.5713 ± 0.0490	0.0786 ± 0.0109
	UMAP	KNN	0.0140 ± 0.0030	1.7963 ± 0.2343
		MLP	2.9829 ± 0.5787	1.9269 ± 0.2816
		SVM	2.8639 ± 0.2055	1.7045 ± 0.1684
		RF	9.7846 ± 0.3915	3.2700 ± 0.2462
	Ivis	KNN	134.9632 ± 56.828	0.6782 ± 0.1825
		MLP	36.1116 ± 8.3472	0.6553 ± 0.1514
		SVM	120.9034 ± 43.9869	0.4351 ± 0.0936
		RF	53.2667 ± 6.8982	0.7636 ± 0.1581
28	PCA	KNN	0.0099 ± 0.0014	0.0079 ± 0.0004
		MLP	2.1493 ± 0.1470	0.0024 ± 0.0009
		SVM	0.2459 ± 0.0623	0.0014 ± 0.0001
		RF	0.2214 ± 0.0172	0.0335 ± 0.0118

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0083 ± 0.0026	0.0070 ± 0.0003
		MLP	0.6281 ± 0.0411	0.0022 ± 0.0006
		SVM	0.1675 ± 0.0403	0.0010 ± 0.0001
		RF	0.2255 ± 0.0162	0.0321 ± 0.0090
	KPCA	KNN	0.0081 ± 0.0007	0.0300 ± 0.0092
		MLP	0.5949 ± 0.0447	0.0754 ± 0.0300
		SVM	0.5853 ± 0.0484	0.0518 ± 0.0212
		RF	0.5932 ± 0.0418	0.0785 ± 0.0150
	UMAP	KNN	10.476 ± 0.3393	4.3785 ± 0.3329
		MLP	4.3977 ± 0.3190	2.6732 ± 0.4316
		SVM	1.7132 ± 0.1333	1.5617 ± 0.2160
		RF	3.5967 ± 0.2141	2.3837 ± 0.2050
	Ivis	KNN	2.8916 ± 0.5912	0.6352 ± 0.2038
		MLP	369.7872 ± 108.5598	0.6724 ± 0.3168
		SVM	268.4045 ± 59.5753	0.9055 ± 0.1387
		RF	76.4436 ± 19.9328	0.4515 ± 0.2516
29	PCA	KNN	0.0118 ± 0.0051	0.0072 ± 0.0005
		MLP	0.8662 ± 0.0439	0.0028 ± 0.0004
		SVM	0.0362 ± 0.0025	0.0016 ± 0.0003
		RF	0.3839 ± 0.0299	0.0451 ± 0.0155
	TSVD	KNN	0.0074 ± 0.0016	0.0067 ± 0.0005
		MLP	3.4007 ± 0.6135	0.0020 ± 0.0000
		SVM	0.0292 ± 0.0038	0.0016 ± 0.0004
		RF	0.2278 ± 0.0275	0.0384 ± 0.0221
	KPCA	KNN	0.0079 ± 0.0033	0.0439 ± 0.0099
		MLP	3.1979 ± 0.1817	0.1266 ± 0.0546
		SVM	0.1958 ± 0.0614	0.0433 ± 0.0245
		RF	0.7309 ± 0.0349	0.0858 ± 0.0092
	UMAP	KNN	3.9188 ± 0.2898	1.7076 ± 0.2201
		MLP	0.3333 ± 0.0136	4.5764 ± 0.3740

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	1.9379 ± 0.2045	1.6420 ± 0.2191
		RF	6.3215 ± 0.1962	4.1373 ± 0.3801
	Ivis	KNN	152.0078 ± 40.5623	0.7027 ± 0.2352
		MLP	61.9691 ± 25.5109	0.6329 ± 0.2065
		SVM	131.9114 ± 33.2568	0.4883 ± 0.2305
		RF	1.9629 ± 0.3414	0.4677 ± 0.1551
30	PCA	KNN	0.0103 ± 0.0017	0.0075 ± 0.0002
		MLP	0.3886 ± 0.0217	0.0019 ± 0.0003
		SVM	0.0284 ± 0.0067	0.0018 ± 0.0007
		RF	0.3690 ± 0.0232	0.0311 ± 0.0060
	TSVD	KNN	0.0071 ± 0.0009	0.0033 ± 0.0003
		MLP	1.6658 ± 0.1320	0.0018 ± 0.0005
		SVM	0.0319 ± 0.0042	0.0018 ± 0.0006
		RF	0.6306 ± 0.0297	0.0517 ± 0.0129
	KPCA	KNN	0.0747 ± 0.0182	0.0388 ± 0.0101
		MLP	1.4517 ± 0.1232	0.0669 ± 0.0283
		SVM	0.1612 ± 0.0249	0.0375 ± 0.0290
		RF	0.5945 ± 0.0311	0.0762 ± 0.0106
	UMAP	KNN	2.6190 ± 0.2574	1.6599 ± 0.1558
		MLP	1.7585 ± 0.0984	1.6348 ± 0.3000
		SVM	2.7962 ± 0.1692	2.0067 ± 0.1604
		RF	3.8396 ± 0.2011	2.5074 ± 0.1715
	Ivis	KNN	333.75 ± 116.4168	0.5788 ± 0.4846
		MLP	119.5841 ± 31.7542	0.7092 ± 0.1412
		SVM	86.9161 ± 26.674	0.3758 ± 0.1451
		RF	28.1767 ± 5.3160	0.5102 ± 0.2807
31	PCA	KNN	0.0091 ± 0.0016	0.0022 ± 0.0003
		MLP	1.3740 ± 0.0721	0.0020 ± 0.0002
		SVM	0.0299 ± 0.0092	0.0018 ± 0.0004
		RF	0.6520 ± 0.0422	0.0578 ± 0.0155

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0067 ± 0.0006	0.0075 ± 0.0005
		MLP	5.3748 ± 1.7997	0.0015 ± 0.0003
		SVM	0.2899 ± 0.0516	0.0011 ± 0.0002
		RF	0.5668 ± 0.0405	0.0531 ± 0.0118
	KPCA	KNN	0.0673 ± 0.0169	0.0370 ± 0.0139
		MLP	1.5476 ± 0.0829	0.0467 ± 0.0311
		SVM	0.2116 ± 0.0623	0.0413 ± 0.0198
		RF	0.6866 ± 0.0323	0.0979 ± 0.0115
	UMAP	KNN	9.3902 ± 0.3022	1.7160 ± 0.2255
		MLP	3.5069 ± 0.4238	1.8825 ± 0.1821
		SVM	10.7489 ± 0.3771	3.0829 ± 0.3995
		RF	9.7167 ± 0.3042	3.7724 ± 0.3334
	Ivis	KNN	2.0627 ± 0.3936	0.6273 ± 0.1709
		MLP	230.8176 ± 40.7631	0.6757 ± 0.1933
		SVM	1.1098 ± 0.3185	0.5761 ± 0.1310
		RF	110.1333 ± 45.6133	1.0018 ± 0.2887
32	PCA	KNN	0.0122 ± 0.0047	0.0073 ± 0.0005
		MLP	0.8505 ± 0.0819	0.0022 ± 0.0004
		SVM	0.4837 ± 0.0866	0.0015 ± 0.0002
		RF	0.2172 ± 0.0268	0.0227 ± 0.0127
	TSVD	KNN	0.0076 ± 0.0014	0.0029 ± 0.0004
		MLP	2.6015 ± 0.2041	0.0020 ± 0.0001
		SVM	0.4584 ± 0.0912	0.0014 ± 0.0002
		RF	0.5495 ± 0.0287	0.0516 ± 0.0152
	KPCA	KNN	0.0721 ± 0.0124	0.0297 ± 0.0065
		MLP	2.1723 ± 0.0576	0.0419 ± 0.0157
		SVM	0.0968 ± 0.0207	0.0314 ± 0.0129
		RF	0.3646 ± 0.0269	0.0467 ± 0.0048
	UMAP	KNN	3.0873 ± 0.2861	2.9447 ± 0.2722
		MLP	13.8362 ± 0.6468	3.5511 ± 0.4260

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	1.9756 ± 0.2233	1.7121 ± 0.2706
		RF	2.2622 ± 0.1368	1.6540 ± 0.1525
	Ivis	KNN	240.7506 ± 49.630	0.6643 ± 0.2098
		MLP	42.0049 ± 7.8658	0.5829 ± 0.1399
		SVM	1.1825 ± 0.5025	0.4858 ± 0.3647
		RF	125.9971 ± 43.741	1.0155 ± 0.1931
33	PCA	KNN	0.0100 ± 0.0025	0.0101 ± 0.0037
		MLP	1.5571 ± 0.1062	0.0018 ± 0.0003
		SVM	0.2071 ± 0.0474	0.0015 ± 0.0003
		RF	0.6489 ± 0.0436	0.0532 ± 0.0163
	TSVD	KNN	0.0070 ± 0.0010	0.0027 ± 0.0011
		MLP	3.1614 ± 0.1162	0.0021 ± 0.0001
		SVM	0.3946 ± 0.0667	0.0015 ± 0.0001
		RF	0.2371 ± 0.0113	0.0258 ± 0.0056
	KPCA	KNN	0.0850 ± 0.0135	0.0330 ± 0.0067
		MLP	0.8910 ± 0.0625	0.0738 ± 0.0215
		SVM	0.3857 ± 0.0786	0.0261 ± 0.0128
		RF	0.6670 ± 0.0310	0.0758 ± 0.0138
	UMAP	KNN	2.6656 ± 0.2347	1.9349 ± 0.2423
		MLP	2.8556 ± 0.1633	4.6950 ± 0.2321
		SVM	3.6445 ± 0.1830	1.5673 ± 0.1624
		RF	2.2098 ± 0.2134	1.7150 ± 0.1525
	Ivis	KNN	120.4023 ± 39.0744	0.5391 ± 0.2047
		MLP	17.6579 ± 2.9529	0.3744 ± 0.1087
		SVM	1.7456 ± 0.1981	0.8940 ± 0.1773
		RF	126.1532 ± 20.5833	1.2101 ± 0.2019
34	PCA	KNN	0.0091 ± 0.0037	0.0030 ± 0.0006
		MLP	1.2375 ± 0.0958	0.0018 ± 0.0000
		SVM	0.3668 ± 0.0737	0.0016 ± 0.0001
		RF	0.2199 ± 0.0195	0.0357 ± 0.0115

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0082 ± 0.0008	0.0124 ± 0.0086
		MLP	2.2704 ± 0.0999	0.0020 ± 0.0002
		SVM	0.4410 ± 0.1002	0.0016 ± 0.0001
		RF	0.5006 ± 0.0471	0.0329 ± 0.0051
	KPCA	KNN	0.0110 ± 0.0033	0.0460 ± 0.0118
		MLP	0.1842 ± 0.0290	0.0612 ± 0.0193
		SVM	0.1116 ± 0.0206	0.0503 ± 0.0341
		RF	0.3997 ± 0.0193	0.0855 ± 0.0275
	UMAP	KNN	3.3390 ± 0.2234	1.6388 ± 0.1892
		MLP	2.4376 ± 0.2239	3.6340 ± 0.3894
		SVM	2.0567 ± 0.1686	2.8480 ± 0.1116
		RF	3.9596 ± 0.4332	1.1798 ± 0.1188
	Ivis	KNN	34.9966 ± 7.4713	0.2354 ± 0.0471
		MLP	9.8450 ± 2.7992	0.6146 ± 0.4946
		SVM	130.1573 ± 34.6766	0.4359 ± 0.1165
		RF	218.0682 ± 60.836	1.4600 ± 0.4110
35	PCA	KNN	0.0084 ± 0.0010	0.0032 ± 0.0004
		MLP	3.1197 ± 0.2243	0.0022 ± 0.0001
		SVM	0.2161 ± 0.0510	0.0019 ± 0.0011
		RF	0.6707 ± 0.1041	0.0521 ± 0.0068
	TSVD	KNN	0.0095 ± 0.0044	0.0045 ± 0.0050
		MLP	1.4946 ± 0.1384	0.0020 ± 0.0001
		SVM	0.1220 ± 0.0452	0.0016 ± 0.0002
		RF	0.5410 ± 0.0259	0.0342 ± 0.0035
	KPCA	KNN	0.0756 ± 0.0154	0.0440 ± 0.0155
		MLP	0.7328 ± 0.0559	0.1253 ± 0.0821
		SVM	0.5401 ± 0.0591	0.0468 ± 0.0245
		RF	0.6073 ± 0.0524	0.0825 ± 0.0119
	UMAP	KNN	3.8346 ± 0.2112	1.7053 ± 0.2349
		MLP	0.4103 ± 0.0866	5.6652 ± 0.3418

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	2.7727 ± 0.2413	1.6013 ± 0.1816
		RF	7.5182 ± 0.3217	2.5723 ± 0.1781
	Ivis	KNN	163.1738 ± 49.029	0.6810 ± 0.2467
		MLP	124.3543 ± 26.4214	0.6087 ± 0.1997
		SVM	127.8655 ± 54.2593	0.7104 ± 0.2307
		RF	405.3471 ± 84.6522	1.4028 ± 0.2708
36	PCA	KNN	0.0093 ± 0.0022	0.0031 ± 0.0003
		MLP	2.2714 ± 0.4031	0.0024 ± 0.0002
		SVM	0.0373 ± 0.0039	0.0020 ± 0.0003
		RF	0.3538 ± 0.0160	0.0368 ± 0.0080
	TSVD	KNN	0.0087 ± 0.0025	0.0086 ± 0.0013
		MLP	2.2290 ± 0.1179	0.0022 ± 0.0002
		SVM	0.4780 ± 0.0960	0.0015 ± 0.0002
		RF	0.5019 ± 0.0302	0.0427 ± 0.0034
	KPCA	KNN	0.0994 ± 0.0207	0.0563 ± 0.0159
		MLP	2.7000 ± 0.0895	0.0691 ± 0.0234
		SVM	0.3992 ± 0.0810	0.0471 ± 0.0113
		RF	0.5341 ± 0.0296	0.0620 ± 0.0091
	UMAP	KNN	9.6427 ± 0.5025	5.2428 ± 0.4768
		MLP	4.0133 ± 0.3948	1.8559 ± 0.4240
		SVM	4.1477 ± 0.3571	1.7192 ± 0.1708
		RF	4.5980 ± 0.2520	1.8923 ± 0.2091
	Ivis	KNN	55.0086 ± 19.9521	0.7539 ± 0.2047
		MLP	269.318 ± 78.6863	0.4803 ± 0.1017
		SVM	211.5314 ± 64.9904	0.5274 ± 0.1777
		RF	32.6859 ± 6.0400	0.3132 ± 0.1487
37	PCA	KNN	0.0079 ± 0.0008	0.0096 ± 0.0039
		MLP	0.4081 ± 0.0453	0.0019 ± 0.0002
		SVM	0.0345 ± 0.0035	0.0020 ± 0.0001
		RF	0.5942 ± 0.0170	0.0590 ± 0.0188

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0088 ± 0.0031	0.0107 ± 0.0038
		MLP	1.0580 ± 0.1077	0.0023 ± 0.0009
		SVM	0.2489 ± 0.0681	0.0012 ± 0.0003
		RF	0.6156 ± 0.0331	0.0519 ± 0.0146
	KPCA	KNN	0.0740 ± 0.0098	0.0307 ± 0.0084
		MLP	1.8194 ± 0.1563	0.0702 ± 0.0664
		SVM	0.5054 ± 0.0864	0.0421 ± 0.0269
		RF	0.6204 ± 0.0353	0.0811 ± 0.0268
	UMAP	KNN	3.3450 ± 0.1928	4.8412 ± 0.4155
		MLP	10.7409 ± 0.5188	3.8200 ± 0.2573
		SVM	4.7913 ± 0.2899	2.0064 ± 0.1867
		RF	2.9964 ± 0.1855	2.2422 ± 0.2598
	Ivis	KNN	46.4536 ± 15.1133	0.4886 ± 0.1797
		MLP	30.8385 ± 5.1191	0.5785 ± 0.1428
		SVM	37.5148 ± 6.3084	0.4119 ± 0.1321
		RF	82.9754 ± 18.1592	1.0351 ± 0.1633
38	PCA	KNN	0.0090 ± 0.0010	0.0082 ± 0.0005
		MLP	1.5177 ± 0.0989	0.0012 ± 0.0001
		SVM	0.4506 ± 0.1147	0.0016 ± 0.0001
		RF	0.3238 ± 0.0231	0.0397 ± 0.0142
	TSVD	KNN	0.0087 ± 0.0025	0.0076 ± 0.0013
		MLP	1.6625 ± 0.1006	0.0022 ± 0.0002
		SVM	0.1138 ± 0.0406	0.0014 ± 0.0003
		RF	0.4239 ± 0.0332	0.0512 ± 0.0265
	KPCA	KNN	0.0095 ± 0.0015	0.0300 ± 0.0048
		MLP	1.7492 ± 0.1878	0.0187 ± 0.0025
		SVM	0.4673 ± 0.0699	0.0333 ± 0.0203
		RF	0.7224 ± 0.0410	0.0808 ± 0.0086
	UMAP	KNN	11.5423 ± 0.4042	3.2858 ± 0.3635
		MLP	11.3873 ± 0.4287	4.9544 ± 0.5109

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	6.2348 ± 0.3391	2.3555 ± 0.2756
		RF	0.3686 ± 0.0258	1.9655 ± 0.1759
	Ivis	KNN	34.1533 ± 4.4262	0.4138 ± 0.0833
		MLP	69.1904 ± 13.5698	0.2079 ± 0.0301
		SVM	39.8214 ± 4.3002	0.5629 ± 0.1826
		RF	359.81 ± 91.0914	1.3178 ± 0.2565
39	PCA	KNN	0.0089 ± 0.0008	0.0032 ± 0.0005
		MLP	3.4763 ± 0.3132	0.0044 ± 0.0059
		SVM	0.0531 ± 0.0123	0.0020 ± 0.0002
		RF	0.6568 ± 0.0186	0.0568 ± 0.0138
	TSVD	KNN	0.0102 ± 0.0037	0.0117 ± 0.0050
		MLP	3.7970 ± 0.4968	0.0021 ± 0.0001
		SVM	0.0494 ± 0.0143	0.0022 ± 0.0004
		RF	0.3205 ± 0.0211	0.0394 ± 0.0100
	KPCA	KNN	0.0100 ± 0.0014	0.0440 ± 0.0070
		MLP	0.3925 ± 0.0556	0.0215 ± 0.0017
		SVM	0.1303 ± 0.0129	0.0365 ± 0.0092
		RF	0.6419 ± 0.0349	0.0721 ± 0.0219
	UMAP	KNN	3.8541 ± 0.3241	1.6584 ± 0.1476
		MLP	10.9974 ± 0.3426	2.8790 ± 0.2391
		SVM	4.2891 ± 0.2089	1.7115 ± 0.1678
		RF	9.0734 ± 0.2698	2.9663 ± 0.1334
	Ivis	KNN	201.563 ± 33.8661	0.3036 ± 0.1468
		MLP	132.25 ± 19.6785	0.3455 ± 0.1752
		SVM	101.9189 ± 22.4975	0.5520 ± 0.2785
		RF	165.688 ± 23.109	1.0249 ± 0.2990
40	PCA	KNN	0.0096 ± 0.0023	0.0087 ± 0.0003
		MLP	0.4956 ± 0.0386	0.0020 ± 0.0002
		SVM	0.2208 ± 0.0671	0.0015 ± 0.0003
		RF	0.3533 ± 0.0441	0.0433 ± 0.0092

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0088 ± 0.0030	0.0049 ± 0.0055
		MLP	2.2999 ± 0.2109	0.0011 ± 0.0001
		SVM	0.2946 ± 0.1223	0.0016 ± 0.0001
		RF	0.4449 ± 0.0300	0.0376 ± 0.0053
	KPCA	KNN	0.0104 ± 0.0041	0.0345 ± 0.0087
		MLP	1.8197 ± 0.0656	0.0557 ± 0.0320
		SVM	0.3437 ± 0.0947	0.0561 ± 0.0274
		RF	0.5323 ± 0.0543	0.0503 ± 0.0107
	UMAP	KNN	1.9893 ± 0.2540	1.5797 ± 0.1429
		MLP	15.266 ± 0.3539	4.7833 ± 0.4058
		SVM	6.0806 ± 4.4883	7.5774 ± 4.2577
		RF	5.2019 ± 0.1972	5.2755 ± 0.3328
	Ivis	KNN	0.8013 ± 0.2006	0.2865 ± 0.0826
		MLP	1.5948 ± 0.2010	0.3534 ± 0.1545
		SVM	249.2861 ± 80.6602	0.9979 ± 0.1856
		RF	68.8423 ± 17.3899	1.1312 ± 0.2536
41	PCA	KNN	0.0095 ± 0.0016	0.0099 ± 0.0004
		MLP	1.5791 ± 0.1111	0.0020 ± 0.0002
		SVM	0.3659 ± 0.0815	0.0016 ± 0.0001
		RF	0.5520 ± 0.0294	0.0460 ± 0.0084
	TSVD	KNN	0.0072 ± 0.0017	0.0021 ± 0.0004
		MLP	2.3050 ± 0.3473	0.0025 ± 0.0001
		SVM	0.5182 ± 0.0787	0.0015 ± 0.0003
		RF	0.5087 ± 0.0631	0.0501 ± 0.0136
	KPCA	KNN	0.0096 ± 0.0019	0.0403 ± 0.0079
		MLP	4.6511 ± 0.3206	0.0644 ± 0.0720
		SVM	0.5489 ± 0.0934	0.0422 ± 0.0258
		RF	0.4547 ± 0.0434	0.0650 ± 0.0065
	UMAP	KNN	2.3390 ± 0.2323	1.5799 ± 0.1491
		MLP	13.5633 ± 0.3111	2.7179 ± 0.1946

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	4.0438 ± 0.2027	1.7316 ± 0.1753
		RF	4.7630 ± 0.2529	1.8180 ± 0.1749
	Ivis	KNN	74.2343 ± 23.2281	0.6307 ± 0.2407
		MLP	4.1157 ± 1.0867	0.6533 ± 0.2680
		SVM	69.6339 ± 20.3423	0.3314 ± 0.0942
		RF	150.5798 ± 45.8041	1.1897 ± 0.2973
42	PCA	KNN	0.0101 ± 0.0032	0.0138 ± 0.0064
		MLP	2.5888 ± 0.1287	0.0026 ± 0.0002
		SVM	0.0565 ± 0.0062	0.0023 ± 0.0010
		RF	0.5838 ± 0.0226	0.0635 ± 0.0146
	TSVD	KNN	0.0105 ± 0.0037	0.0117 ± 0.0027
		MLP	1.7651 ± 0.1238	0.0018 ± 0.0004
		SVM	0.0573 ± 0.0075	0.0031 ± 0.0032
		RF	0.6409 ± 0.0519	0.0497 ± 0.0095
	KPCA	KNN	0.0585 ± 0.0101	0.0270 ± 0.0065
		MLP	2.4763 ± 0.0950	0.1002 ± 0.1061
		SVM	0.1136 ± 0.0179	0.0493 ± 0.0232
		RF	0.5906 ± 0.0670	0.0986 ± 0.0955
	UMAP	KNN	3.4007 ± 0.2428	1.8064 ± 0.2784
		MLP	2.4412 ± 0.2005	2.2443 ± 0.2843
		SVM	2.1903 ± 0.1577	2.5140 ± 0.2571
		RF	4.7399 ± 0.2777	1.7111 ± 0.1248
	Ivis	KNN	283.2155 ± 39.0904	0.6486 ± 0.2641
		MLP	32.3804 ± 5.6662	0.6817 ± 0.2226
		SVM	37.0951 ± 8.1949	0.7519 ± 0.1987
		RF	91.4552 ± 30.2741	1.0054 ± 0.2319
43	PCA	KNN	0.0088 ± 0.0014	0.0024 ± 0.0002
		MLP	1.9721 ± 0.1468	0.0020 ± 0.0002
		SVM	0.2553 ± 0.0649	0.0021 ± 0.0020
		RF	0.3799 ± 0.0288	0.0390 ± 0.0105

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0105 ± 0.0044	0.0105 ± 0.0009
		MLP	1.9504 ± 0.0684	0.0024 ± 0.0003
		SVM	0.0505 ± 0.0054	0.0022 ± 0.0007
		RF	0.6856 ± 0.0460	0.0557 ± 0.0136
	KPCA	KNN	0.0767 ± 0.0128	0.0496 ± 0.0131
		MLP	1.6324 ± 0.0713	0.1115 ± 0.0546
		SVM	0.0385 ± 0.0146	0.0384 ± 0.0106
		RF	0.7184 ± 0.0368	0.0946 ± 0.0218
	UMAP	KNN	3.1633 ± 0.1738	1.6564 ± 0.2082
		MLP	8.5027 ± 0.4582	1.2726 ± 0.1519
		SVM	3.3713 ± 0.1527	1.6687 ± 0.2126
		RF	2.5952 ± 0.1480	1.7640 ± 0.1938
	Ivis	KNN	168.4867 ± 93.2676	0.5938 ± 0.1354
		MLP	36.6792 ± 6.0608	0.6692 ± 0.1953
		SVM	61.7638 ± 30.8461	0.3806 ± 0.1634
		RF	85.1242 ± 19.0671	1.2539 ± 0.2720
44	PCA	KNN	0.0090 ± 0.0011	0.0094 ± 0.0002
		MLP	1.7702 ± 0.1151	0.0024 ± 0.0002
		SVM	0.2395 ± 0.0416	0.0013 ± 0.0004
		RF	0.6251 ± 0.0348	0.0561 ± 0.0162
	TSVD	KNN	0.0102 ± 0.0011	0.0027 ± 0.0006
		MLP	0.3848 ± 0.0429	0.0020 ± 0.0001
		SVM	0.5239 ± 0.0981	0.0026 ± 0.0024
		RF	0.3582 ± 0.0186	0.0438 ± 0.0131
	KPCA	KNN	0.0653 ± 0.0130	0.0479 ± 0.0161
		MLP	1.2333 ± 0.0937	0.0788 ± 0.0310
		SVM	0.4430 ± 0.0705	0.0522 ± 0.0296
		RF	0.6947 ± 0.0504	0.1051 ± 0.0309
	UMAP	KNN	1.7049 ± 0.1827	1.5645 ± 0.1464
		MLP	14.2031 ± 0.7122	5.6882 ± 0.6146

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	8.8395 ± 0.3801	1.9022 ± 0.3491
		RF	3.0732 ± 0.1645	1.6239 ± 0.1687
	Ivis	KNN	1.5970 ± 0.6701	0.7517 ± 0.4755
		MLP	1.3856 ± 0.3714	0.2220 ± 0.0226
		SVM	0.7697 ± 0.0589	0.2258 ± 0.0411
		RF	83.4254 ± 11.4361	0.3413 ± 0.1305
45	PCA	KNN	0.0081 ± 0.0007	0.0127 ± 0.0048
		MLP	0.4544 ± 0.0360	0.0019 ± 0.0003
		SVM	0.0535 ± 0.0045	0.0016 ± 0.0003
		RF	0.6518 ± 0.0442	0.0518 ± 0.0080
	TSVD	KNN	0.0084 ± 0.0015	0.0102 ± 0.0006
		MLP	1.3643 ± 0.1003	0.0019 ± 0.0001
		SVM	0.1104 ± 0.0277	0.0017 ± 0.0003
		RF	0.6290 ± 0.0365	0.0392 ± 0.0083
	KPCA	KNN	0.0818 ± 0.0104	0.0437 ± 0.0118
		MLP	4.2740 ± 0.4390	0.0145 ± 0.0017
		SVM	0.3798 ± 0.0783	0.0515 ± 0.0150
		RF	0.3491 ± 0.0236	0.0461 ± 0.0060
	UMAP	KNN	9.0921 ± 0.3533	2.3763 ± 0.2442
		MLP	4.3703 ± 2.5366	2.0864 ± 1.3167
		SVM	8.1197 ± 0.2486	5.5723 ± 0.7266
		RF	10.5217 ± 0.3440	2.2898 ± 0.1339
	Ivis	KNN	36.4105 ± 8.3548	0.5327 ± 0.1324
		MLP	262.1692 ± 40.8057	0.3207 ± 0.1851
		SVM	75.7727 ± 13.0771	0.6088 ± 0.1462
		RF	60.1523 ± 11.3307	0.6398 ± 0.1839
46	PCA	KNN	0.0095 ± 0.0015	0.0031 ± 0.0004
		MLP	0.3202 ± 0.0711	0.0018 ± 0.0002
		SVM	0.1047 ± 0.0150	0.0017 ± 0.0001
		RF	0.3484 ± 0.0188	0.0340 ± 0.0058

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0090 ± 0.0023	0.0022 ± 0.0004
		MLP	0.4353 ± 0.0183	0.0019 ± 0.0002
		SVM	0.3138 ± 0.0711	0.0017 ± 0.0001
		RF	0.2931 ± 0.0199	0.0307 ± 0.0074
	KPCA	KNN	0.0828 ± 0.0092	0.0489 ± 0.0112
		MLP	0.1012 ± 0.0158	0.0609 ± 0.0229
		SVM	0.1346 ± 0.0295	0.0385 ± 0.0089
		RF	0.2305 ± 0.0296	0.0454 ± 0.0044
	UMAP	KNN	1.8308 ± 0.2175	1.4607 ± 0.1885
		MLP	7.4858 ± 0.1919	2.0343 ± 0.2735
		SVM	16.9492 ± 0.5409	5.9715 ± 0.5034
		RF	2.2488 ± 0.2244	1.5462 ± 0.1666
	Ivis	KNN	171.664 ± 61.4779	0.7839 ± 0.2052
		MLP	234.4598 ± 64.068	0.4834 ± 0.2179
		SVM	264.9783 ± 43.7364	0.5873 ± 0.1993
		RF	33.2403 ± 7.2009	0.5962 ± 0.2291
47	PCA	KNN	0.0103 ± 0.0027	0.0126 ± 0.0059
		MLP	0.6993 ± 0.0271	0.0025 ± 0.0008
		SVM	0.0409 ± 0.0051	0.0024 ± 0.0010
		RF	0.5862 ± 0.0437	0.0554 ± 0.0151
	TSVD	KNN	0.0082 ± 0.0008	0.0111 ± 0.0004
		MLP	1.1645 ± 0.0937	0.0024 ± 0.0002
		SVM	0.3684 ± 0.0495	0.0016 ± 0.0001
		RF	0.6731 ± 0.0359	0.0622 ± 0.0143
	KPCA	KNN	0.0750 ± 0.0134	0.0511 ± 0.0148
		MLP	2.3824 ± 0.6037	0.0574 ± 0.0207
		SVM	0.6854 ± 0.1056	0.0531 ± 0.0226
		RF	0.7335 ± 0.0479	0.0856 ± 0.0100
	UMAP	KNN	4.1825 ± 0.2110	1.8593 ± 0.1997
		MLP	2.6768 ± 0.2692	3.0687 ± 0.1939

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
		SVM	3.1675 ± 0.2608	1.7191 ± 0.1853
		RF	1.9420 ± 0.1629	1.6397 ± 0.1519
	Ivis	KNN	1.6055 ± 0.4218	0.8577 ± 0.4154
		MLP	6.2063 ± 0.5963	0.5552 ± 0.1725
		SVM	90.9303 ± 27.3378	0.2316 ± 0.1144
		RF	35.5297 ± 13.7196	0.4411 ± 0.1746
48	PCA	KNN	0.0090 ± 0.0012	0.0027 ± 0.0005
		MLP	0.6313 ± 0.0981	0.0033 ± 0.0037
		SVM	0.3284 ± 0.0770	0.0017 ± 0.0002
		RF	0.4726 ± 0.0236	0.0522 ± 0.0176
	TSVD	KNN	0.0080 ± 0.0015	0.0143 ± 0.0064
		MLP	0.6086 ± 0.0549	0.0023 ± 0.0003
		SVM	0.2151 ± 0.0490	0.0011 ± 0.0003
		RF	0.2246 ± 0.0164	0.0344 ± 0.0090
	KPCA	KNN	0.0669 ± 0.0125	0.0414 ± 0.0133
		MLP	1.9509 ± 0.1067	0.0733 ± 0.0213
		SVM	0.2149 ± 0.0485	0.0149 ± 0.0029
		RF	0.6109 ± 0.0682	0.0562 ± 0.0112
	UMAP	KNN	0.0127 ± 0.0034	1.9223 ± 0.2548
		MLP	4.6520 ± 0.2839	6.1008 ± 0.3630
		SVM	6.0742 ± 0.2930	1.8209 ± 0.2061
		RF	15.1064 ± 0.4097	6.1525 ± 0.3879
	Ivis	KNN	79.7708 ± 8.9398	0.3241 ± 0.2006
		MLP	0.8744 ± 0.1358	0.2471 ± 0.1280
		SVM	37.7916 ± 6.7825	0.4380 ± 0.1605
		RF	50.7492 ± 10.4672	0.8414 ± 0.1984
49	PCA	KNN	0.0120 ± 0.0038	0.0101 ± 0.0004
		MLP	0.9028 ± 0.0493	0.0024 ± 0.0003
		SVM	0.0513 ± 0.0070	0.0026 ± 0.0004
		RF	0.3720 ± 0.0296	0.0672 ± 0.0967

Continued on next page

Table 3 – Mean execution time of best-performing pipelines in 10-fold stratified cross-validation.

		Measure	Fit time	Score time
Dimensionality	Projector	Classifier		
	TSVD	KNN	0.0109 ± 0.0052	0.0111 ± 0.0019
		MLP	1.5434 ± 0.1238	0.0019 ± 0.0001
		SVM	0.0323 ± 0.0018	0.0020 ± 0.0002
		RF	0.4863 ± 0.0241	0.0384 ± 0.0060
	KPCA	KNN	0.0865 ± 0.0109	0.0342 ± 0.0097
		MLP	1.6898 ± 0.1389	0.0978 ± 0.0630
		SVM	0.0402 ± 0.0141	0.0348 ± 0.0142
		RF	0.3022 ± 0.0403	0.0547 ± 0.0125
	UMAP	KNN	2.4426 ± 0.2456	6.3907 ± 0.4150
		MLP	4.3421 ± 0.3050	0.0000 ± 0.0000
		SVM	5.8680 ± 0.2446	2.0936 ± 0.2203
		RF	4.6136 ± 0.1365	1.9876 ± 0.1772
	Ivis	KNN	49.9125 ± 64.750	0.6607 ± 0.1412
		MLP	303.6793 ± 70.6738	0.7313 ± 0.2597
		SVM	116.3505 ± 37.6608	0.6545 ± 0.2303
		RF	104.6235 ± 32.1944	1.0320 ± 0.3296

APPENDIX

C

BEST HYPER-PARAMETERS FOUND

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
50	—	Projector	—	—	—	—
		Classifier	leaf_size: 26 n_neighbors: 3 p: 1 weights: distance	activation: relu alpha: 0.7215564227807617 layer_one: 67 layer_two: 54 max_iter: 2000	C: 20.41617597928149 gamma: 68.90834971976716 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 50 n_jobs: -1
2	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 25 p: 5 weights: uniform	activation: relu alpha: 0.18737835928036276 layer_one: 82 layer_two: 124 max_iter: 1194	C: 26.80380173673879 gamma: 48.08905220445172 kernel: linear max_iter: 100000 probability: True	max_depth: 3 n_estimators: 135 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack activation: tanh	algorithm: arpack C: 7.1205395004936	algorithm: arpack
		Classifier	leaf_size: 17 n_neighbors: 25 p: 1 weights: distance	alpha: 0.0001 layer_one: 85 layer_two: 106 max_iter: 479	gamma: 40.62747792192799 kernel: linear max_iter: 100000 probability: True	max_depth: 12 n_estimators: 113 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.7631660469278523 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.18695442421512531 kernel: linear n_jobs: -1 activation: relu	eigen_solver: dense gamma: 0.5129059051480938 kernel: linear n_jobs: -1 C: 127.65749874708352	eigen_solver: dense gamma: 0.1168052744110393 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 21 n_neighbors: 16 p: 4 weights: distance	alpha: 0.0001 layer_one: 90 layer_two: 50 max_iter: 2000	gamma: 64.8744744583126 kernel: linear max_iter: 100000 probability: True	max_depth: 11 n_estimators: 50 n_jobs: -1
	UMAP	Projector	n_epochs: 1694 n_neighbors: 3 target_weight: 0.6887647298354286	n_epochs: 282 n_neighbors: 15 target_weight: 0.6916103701632366 activation: tanh	n_epochs: 2000 n_neighbors: 3 target_weight: 0.2387735663333006 C: 985.2902676692498	n_epochs: 2000 n_neighbors: 3 target_weight: 0.7879949317192984
		Classifier	leaf_size: 15 n_neighbors: 4 p: 3 weights: distance	alpha: 0.0001 layer_one: 69 layer_two: 50 max_iter: 100	gamma: 56.59585546539506 kernel: rbf max_iter: 100000 probability: True	max_depth: 3 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.8013832491892193	k: 38 model: hinton n_epochs_without_progress: 40 supervision_weight: 1.0 activation: tanh	k: 3 model: hinton n_epochs_without_progress: 30 supervision_weight: 1.0 C: 447.8756201510728	k: 29 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0
		Classifier	leaf_size: 3 n_neighbors: 30 p: 1 weights: distance	alpha: 0.0001 layer_one: 127 layer_two: 150 max_iter: 2000	gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 14 n_estimators: 150 n_jobs: -1
3	PCA	Projector	—	—	—	—

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 29 n_neighbors: 19 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 55 layer_two: 50 max_iter: 100	C: 331.3969277503116 gamma: 6.830188552360055 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 62 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 4 n_neighbors: 19 p: 3 weights: distance	activation: relu alpha: 0.6779758749165763 layer_one: 150 layer_two: 50 max_iter: 100	C: 629.9676083781482 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 127 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.6777859830649108 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.19497722737592268 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.31690871193087605 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: linear n_jobs: -1
		Classifier	leaf_size: 29 n_neighbors: 17 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 82 layer_two: 50 max_iter: 2000	C: 794.869407756802 gamma: 0.01 kernel: sigmoid max_iter: 100000 probability: True	max_depth: 8 n_estimators: 50 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0	n_epochs: 1852 n_neighbors: 25 target_weight: 0.0	n_epochs: 960 n_neighbors: 3 target_weight: 0.8473658289906908	n_epochs: 1987 n_neighbors: 4 target_weight: 0.0
		Classifier	leaf_size: 6 n_neighbors: 3 p: 5 weights: uniform	activation: tanh alpha: 8.46485306798708 layer_one: 126 layer_two: 114 max_iter: 1552	C: 255.06009716561346 gamma: 64.4382973678162 kernel: rbf max_iter: 100000 probability: True	max_depth: 4 n_estimators: 82 n_jobs: -1
	Ivis	Projector	k: 28 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 96 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.9159628556325711	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 3 model: hinton n_epochs_without_progress: 30 supervision_weight: 0.7449338390496758
		Classifier	leaf_size: 5 n_neighbors: 3 p: 2 weights: distance	activation: tanh alpha: 0.3990511961539001 layer_one: 50 layer_two: 68 max_iter: 1240	C: 898.4164364790793 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 134 n_jobs: -1
4	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 7 n_neighbors: 11 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 150 layer_two: 150 max_iter: 750	C: 672.1836386895451 gamma: 21.496427870082346 kernel: rbf max_iter: 100000 probability: True	max_depth: 10 n_estimators: 55 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 6 n_neighbors: 10 p: 1 weights: distance	activation: tanh alpha: 1.0762776822388889 layer_one: 148 layer_two: 147 max_iter: 242	C: 477.02731472656626 gamma: 21.056449632068745 kernel: linear max_iter: 100000 probability: True	max_depth: 9 n_estimators: 83 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.6560783291624676 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.3135993674111598 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.9617682396090604 kernel: linear n_jobs: -1
		Classifier	leaf_size: 22 n_neighbors: 3 p: 3 weights: distance	activation: relu alpha: 4.3999425700454005 layer_one: 124 layer_two: 129 max_iter: 1684	C: 484.6578264318898 gamma: 21.594592685924106 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 77 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.4913588462904141	n_epochs: 2000 n_neighbors: 3 target_weight: 0.332854220712316	n_epochs: 1454 n_neighbors: 3 target_weight: 0.673933785686601	n_epochs: 1784 n_neighbors: 3 target_weight: 0.8110020154119998
		Classifier	leaf_size: 26 n_neighbors: 30 p: 4 weights: distance	activation: relu alpha: 0.0001 layer_one: 121 layer_two: 135 max_iter: 2000	C: 384.4499257022004 gamma: 27.966821800356698 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 71 n_jobs: -1
	Ivis	Projector	k: 91 model: hinton n_epochs_without_progress: 37 supervision_weight: 0.7930974240992923	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7280961228280134	k: 3 model: maaten n_epochs_without_progress: 50 supervision_weight: 0.7603818039884349	k: 5 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.47496347761730734
		Classifier	leaf_size: 30 n_neighbors: 20 p: 5 weights: distance	activation: tanh alpha: 8.021891068618327 layer_one: 150 layer_two: 150 max_iter: 1976	C: 741.1767557028485 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 11 n_estimators: 119 n_jobs: -1
5	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 7 p: 2 weights: distance	activation: relu alpha: 0.3128806946237442 layer_one: 150 layer_two: 50 max_iter: 2000	C: 29.857338005689154 gamma: 3.882627666618701 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 61 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 4 n_neighbors: 11 p: 1 weights: distance	activation: relu alpha: 0.5735976399312728 layer_one: 131 layer_two: 79 max_iter: 1989	C: 286.0729135542215 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 10 n_estimators: 116 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.5120292154942286 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.37266380761715334 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.7054895297341364 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.04505208541871223 kernel: linear n_jobs: -1

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 29 n_neighbors: 7 p: 2 weights: distance	activation: tanh alpha: 6.965371256602866 layer_one: 119 layer_two: 150 max_iter: 2000	C: 686.1990047454063 gamma: 0.1605810921594878 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 61 n_jobs: -1
	UMAP	Projector	n_epochs: 100 n_neighbors: 3 target_weight: 0.6447661104925521	n_epochs: 1440 n_neighbors: 17 target_weight: 0.929861518430108	n_epochs: 2000 n_neighbors: 3 target_weight: 0.4031102783633555	n_epochs: 2000 n_neighbors: 3 target_weight: 0.13823895111733167
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: distance	activation: tanh alpha: 19.89560714320113 layer_one: 131 layer_two: 75 max_iter: 1776	C: 267.49720370301316 gamma: 100.0 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 3 model: maaten n_epochs_without_progress: 50 supervision_weight: 0.919379660035281	k: 48 model: hinton n_epochs_without_progress: 29 supervision_weight: 1.0	k: 93 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.8683194095074311	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0
		Classifier	leaf_size: 5 n_neighbors: 3 p: 1 weights: distance	activation: relu alpha: 2.9930741497724744 layer_one: 125 layer_two: 50 max_iter: 1660	C: 383.2903258351037 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 50 n_jobs: -1
6	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 5 n_neighbors: 20 p: 1 weights: distance	activation: relu alpha: 0.7118296216873162 layer_one: 50 layer_two: 149 max_iter: 449	C: 969.2940905369417 gamma: 0.4377634833686256 kernel: rbf max_iter: 100000 probability: True	max_depth: 12 n_estimators: 50 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 30 n_neighbors: 10 p: 1 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 150 max_iter: 100	C: 7.541780834286024 gamma: 43.29630392741177 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 76 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.29623003511726703 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.5434458241508222 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.006342493476215427 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.6525341512810248 kernel: linear n_jobs: -1
		Classifier	leaf_size: 25 n_neighbors: 13 p: 2 weights: distance	activation: tanh alpha: 0.0001 layer_one: 50 layer_two: 150 max_iter: 2000	C: 458.2160654618528 gamma: 1.0668174228810572 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 50 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0	n_epochs: 100 n_neighbors: 13 target_weight: 0.9690738287548445	n_epochs: 1697 n_neighbors: 3 target_weight: 0.8029112697362751	n_epochs: 2000 n_neighbors: 3 target_weight: 0.8721618306402672

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: distance	activation: relu alpha: 12.528796300453992 layer_one: 58 layer_two: 107 max_iter: 296	C: 225.3777187156418 gamma: 23.48655340944913 kernel: rbf max_iter: 100000 probability: True	max_depth: 3 n_estimators: 150 n_jobs: -1
	Ivis	Projector	k: 9 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.8983480077984743	k: 57 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0 activation: relu	k: 103 model: hinton n_epochs_without_progress: 42 supervision_weight: 1.0 C: 992.3615319702304	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.8693187535226273
		Classifier	leaf_size: 12 n_neighbors: 3 p: 4 weights: uniform	alpha: 0.0001 layer_one: 76 layer_two: 150 max_iter: 2000	gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 50 n_jobs: -1
7	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 16 p: 1 weights: distance	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 50 max_iter: 387	C: 1000.0 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 50 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack activation: tanh	algorithm: arpack C: 640.5038126157293	algorithm: arpack
		Classifier	leaf_size: 4 n_neighbors: 14 p: 1 weights: distance	alpha: 3.538706800153832 layer_one: 102 layer_two: 85 max_iter: 1969	gamma: 1.068707716638815 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 104 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 1.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: cosine n_jobs: -1 activation: relu	eigen_solver: dense gamma: 0.5318834346334941 kernel: linear n_jobs: -1 C: 536.5782621129622	eigen_solver: dense gamma: 0.0 kernel: linear n_jobs: -1
		Classifier	leaf_size: 27 n_neighbors: 3 p: 4 weights: uniform	alpha: 0.0001 layer_one: 80 layer_two: 150 max_iter: 1880	gamma: 4.393270040962165 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 50 n_jobs: -1
	UMAP	Projector	n_epochs: 460 n_neighbors: 3 target_weight: 0.7824776826777201	n_epochs: 952 n_neighbors: 3 target_weight: 0.38574253404569736 activation: relu	n_epochs: 304 n_neighbors: 3 target_weight: 0.2545922472840518 C: 118.96903686695418	n_epochs: 100 n_neighbors: 3 target_weight: 0.6836549271526163
		Classifier	leaf_size: 29 n_neighbors: 29 p: 4 weights: distance	alpha: 0.0001 layer_one: 59 layer_two: 73 max_iter: 354	gamma: 37.352115446774874 kernel: rbf max_iter: 100000 probability: True	max_depth: 7 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 3 model: hinton n_epochs_without_progress: 35 supervision_weight: 1.0	k: 24 model: hinton n_epochs_without_progress: 17 supervision_weight: 0.897817717546182	k: 65 model: hinton n_epochs_without_progress: 30 supervision_weight: 1.0	k: 3 model: hinton n_epochs_without_progress: 16 supervision_weight: 0.3737271617955752

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 18 p: 1 weights: uniform	activation: relu alpha: 1.5858048515810859 layer_one: 78 layer_two: 50 max_iter: 2000	C: 10.384732062032962 gamma: 15.994427918357168 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 113 n_jobs: -1
8	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 17 p: 1 weights: distance	activation: relu alpha: 0.0001 layer_one: 102 layer_two: 150 max_iter: 100	C: 94.39791905528514 gamma: 2.031803915767622 kernel: rbf max_iter: 100000 probability: True	max_depth: 7 n_estimators: 147 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 27 n_neighbors: 6 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 150 max_iter: 1690	C: 989.5311561874598 gamma: 0.5276002688875562 kernel: rbf max_iter: 100000 probability: True	max_depth: 13 n_estimators: 85 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.8499656176908281 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.6476335738555298 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.44542106776914403 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 0.38629450855090275 kernel: linear n_jobs: -1
		Classifier	leaf_size: 14 n_neighbors: 8 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 97 layer_two: 50 max_iter: 1077	C: 268.1517982377209 gamma: 0.2435312472542982 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 122 n_jobs: -1
	UMAP	Projector	n_epochs: 654 n_neighbors: 3 target_weight: 0.7667851988060171	n_epochs: 1007 n_neighbors: 3 target_weight: 0.820238025999645	n_epochs: 1898 n_neighbors: 3 target_weight: 0.8156402288387831	n_epochs: 100 n_neighbors: 3 target_weight: 0.6841740873319928
		Classifier	leaf_size: 30 n_neighbors: 29 p: 5 weights: distance	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 80 max_iter: 797	C: 1000.0 gamma: 31.56140945644905 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 91 n_jobs: -1
	Ivis	Projector	k: 3 model: hinton n_epochs_without_progress: 30 supervision_weight: 0.752758286898467	k: 3 model: hinton n_epochs_without_progress: 36 supervision_weight: 0.8513240451476765	k: 82 model: hinton n_epochs_without_progress: 43 supervision_weight: 0.9033202621585579	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0
		Classifier	leaf_size: 24 n_neighbors: 3 p: 3 weights: uniform	activation: relu alpha: 0.0001 layer_one: 91 layer_two: 150 max_iter: 2000	C: 874.9981623245496 gamma: 91.36574547438359 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 67 n_jobs: -1
9	PCA	Projector	—	—	—	—

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 15 n_neighbors: 16 p: 1 weights: distance	activation: relu alpha: 1.0904228091598942 layer_one: 120 layer_two: 150 max_iter: 934	C: 384.590638008383 gamma: 0.46045274701218136 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 50 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 4 n_neighbors: 4 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 150 max_iter: 2000	C: 42.679815378757546 gamma: 2.4377019462158933 kernel: rbf max_iter: 100000 probability: True	max_depth: 4 n_estimators: 110 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.6346043423404591 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.8603390351855932 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: linear n_jobs: -1
		Classifier	leaf_size: 17 n_neighbors: 3 p: 1 weights: distance	activation: relu alpha: 0.25038343048638967 layer_one: 50 layer_two: 150 max_iter: 111	C: 924.6423840738752 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 50 n_jobs: -1
	UMAP	Projector	n_epochs: 100 n_neighbors: 3 target_weight: 0.4918849893306672	n_epochs: 400 n_neighbors: 3 target_weight: 0.9226211693511446 activation: tanh	n_epochs: 1234 n_neighbors: 3 target_weight: 0.6461703419514346 C: 158.05924785191127	n_epochs: 1884 n_neighbors: 3 target_weight: 0.24575978206424892
		Classifier	leaf_size: 3 n_neighbors: 30 p: 5 weights: distance	alpha: 0.0001 layer_one: 150 layer_two: 150 max_iter: 1678	gamma: 17.944972187940984 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 150 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 39 supervision_weight: 0.6971520865004183	k: 7 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.40245000844332385 activation: relu	k: 79 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0 C: 998.9825642327704	k: 3 model: maaten n_epochs_without_progress: 42 supervision_weight: 0.5638162769711998
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: distance	alpha: 0.0001 layer_one: 80 layer_two: 111 max_iter: 776	gamma: 57.059481882560384 kernel: linear max_iter: 100000 probability: True	max_depth: 13 n_estimators: 50 n_jobs: -1
10	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 4 n_neighbors: 9 p: 3 weights: distance	activation: relu alpha: 1.2340432565308233 layer_one: 140 layer_two: 150 max_iter: 686	C: 81.37181068845496 gamma: 6.385704833163072 kernel: linear max_iter: 100000 probability: True	max_depth: 8 n_estimators: 85 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 6 n_neighbors: 3 p: 1 weights: distance	activation: relu alpha: 0.0001 layer_one: 94 layer_two: 50 max_iter: 731	C: 766.8422846377921 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 95 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.9975792551082907 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.24554746713760836 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 0.3269046175749575 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.503666643246292 kernel: linear n_jobs: -1
		Classifier	leaf_size: 24 n_neighbors: 3 p: 5 weights: uniform	activation: relu alpha: 0.0001 layer_one: 113 layer_two: 150 max_iter: 160	C: 622.7314909758735 gamma: 0.01 kernel: sigmoid max_iter: 100000 probability: True	max_depth: 8 n_estimators: 91 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.18257017253442542	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0	n_epochs: 1963 n_neighbors: 3 target_weight: 0.5504443587970472	n_epochs: 100 n_neighbors: 3 target_weight: 0.0
		Classifier	leaf_size: 28 n_neighbors: 3 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 150 layer_two: 50 max_iter: 2000	C: 825.4681449400688 gamma: 13.924039964036808 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 79 n_jobs: -1
	Ivis	Projector	k: 21 model: hinton n_epochs_without_progress: 43 supervision_weight: 0.755809554651454	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 62 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 86 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.41325643730259537
		Classifier	leaf_size: 30 n_neighbors: 4 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 50 layer_two: 150 max_iter: 2000	C: 357.08823828234364 gamma: 60.85197618027389 kernel: linear max_iter: 100000 probability: True	max_depth: 14 n_estimators: 150 n_jobs: -1
11	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 6 p: 2 weights: distance	activation: relu alpha: 1.493052261668909 layer_one: 150 layer_two: 134 max_iter: 933	C: 955.3597021386721 gamma: 42.82548637485607 kernel: linear max_iter: 100000 probability: True	max_depth: 7 n_estimators: 150 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: uniform	activation: relu alpha: 0.0001 layer_one: 51 layer_two: 150 max_iter: 2000	C: 895.9505190576311 gamma: 15.454435952581465 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 129 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.6461537752833464 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.9895970349732587 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.8802458655019398 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: linear n_jobs: -1

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 29 n_neighbors: 5 p: 2 weights: distance	activation: relu alpha: 1.1723167024417223 layer_one: 50 layer_two: 70 max_iter: 615	C: 1000.0 gamma: 6.753787887746125 kernel: rbf max_iter: 100000 probability: True	max_depth: 7 n_estimators: 150 n_jobs: -1
	UMAP	Projector	n_epochs: 1761 n_neighbors: 3 target_weight: 0.0	n_epochs: 220 n_neighbors: 3 target_weight: 0.5025084993245902 activation: logistic alpha: 0.0001 layer_one: 103 layer_two: 89 max_iter: 1757	n_epochs: 2000 n_neighbors: 3 target_weight: 0.4301927230842191 C: 933.1182494606477 gamma: 63.729513911148175 kernel: rbf max_iter: 100000 probability: True	n_epochs: 1716 n_neighbors: 3 target_weight: 0.8470506257572668
		Classifier	leaf_size: 30 n_neighbors: 25 p: 1 weights: distance			max_depth: 3 n_estimators: 130 n_jobs: -1
	Ivis	Projector	k: 37 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.6530172395316239	k: 103 model: hinton n_epochs_without_progress: 22 supervision_weight: 1.0 activation: relu alpha: 0.0001 layer_one: 86 layer_two: 131 max_iter: 1978	k: 3 model: hinton n_epochs_without_progress: 47 supervision_weight: 0.8205722535096533 C: 490.19513465795825 gamma: 4.947110685098984 kernel: linear max_iter: 100000 probability: True	k: 83 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.8380175611308899
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: uniform			max_depth: 5 n_estimators: 50 n_jobs: -1
12	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 8 p: 1 weights: distance	activation: relu alpha: 1.0300616459451308 layer_one: 64 layer_two: 50 max_iter: 1443	C: 1000.0 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 8 n_estimators: 110 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 29 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 50 max_iter: 2000	C: 307.535917827501 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 107 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.5863334043765036 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.5807003909662757 kernel: cosine n_jobs: -1 activation: relu alpha: 1.3889050448789841 layer_one: 50 layer_two: 64 max_iter: 1343	eigen_solver: dense gamma: 0.9679313552361881 kernel: cosine n_jobs: -1 C: 265.451899410279 gamma: 7.777837922184481 kernel: rbf max_iter: 100000 probability: True	eigen_solver: dense gamma: 0.9365212809702601 kernel: linear n_jobs: -1
		Classifier	leaf_size: 11 n_neighbors: 8 p: 1 weights: distance			max_depth: 9 n_estimators: 137 n_jobs: -1
	UMAP	Projector	n_epochs: 586 n_neighbors: 3 target_weight: 0.5002148118336228	n_epochs: 2000 n_neighbors: 3 target_weight: 0.8149727878777623	n_epochs: 226 n_neighbors: 3 target_weight: 0.8448263360995929	n_epochs: 1709 n_neighbors: 3 target_weight: 0.0

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 3 p: 4 weights: distance	activation: relu alpha: 0.0001 layer_one: 142 layer_two: 112 max_iter: 1637	C: 122.9319421095117 gamma: 95.18241869848228 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
	Ivis	Projector	k: 78 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7399716000994667	k: 20 model: hinton n_epochs_without_progress: 42 supervision_weight: 0.7531900309040743	k: 23 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 27 model: hinton n_epochs_without_progress: 33 supervision_weight: 0.6520064417592871
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: distance	activation: tanh alpha: 11.792359487651817 layer_one: 78 layer_two: 150 max_iter: 2000	C: 1000.0 gamma: 6.457224005636738 kernel: linear max_iter: 100000 probability: True	max_depth: 10 n_estimators: 50 n_jobs: -1
13	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 9 p: 2 weights: distance	activation: relu alpha: 3.5821731280326334 layer_one: 142 layer_two: 150 max_iter: 1800	C: 163.85219801469233 gamma: 43.4054424898288 kernel: rbf max_iter: 100000 probability: True	max_depth: 11 n_estimators: 114 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 30 n_neighbors: 9 p: 5 weights: distance	activation: relu alpha: 0.7992944119189396 layer_one: 150 layer_two: 68 max_iter: 2000	C: 453.81206803930894 gamma: 17.227808836684883 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 120 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.3070445175001623 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.6751235158015215 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.004616880279978745 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.10732936171444611 kernel: linear n_jobs: -1
		Classifier	leaf_size: 7 n_neighbors: 8 p: 5 weights: distance	activation: relu alpha: 1.2058065005566312 layer_one: 50 layer_two: 50 max_iter: 2000	C: 84.0712066759924 gamma: 10.431649857586324 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 146 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.2858079708854139	n_epochs: 1296 n_neighbors: 3 target_weight: 0.5016231038807054	n_epochs: 539 n_neighbors: 4 target_weight: 0.07961532413437997	n_epochs: 1905 n_neighbors: 3 target_weight: 0.011767146016067324
		Classifier	leaf_size: 3 n_neighbors: 3 p: 5 weights: distance	activation: tanh alpha: 0.0001 layer_one: 148 layer_two: 150 max_iter: 2000	C: 200.77512936457148 gamma: 10.249964872198793 kernel: linear max_iter: 100000 probability: True	max_depth: 6 n_estimators: 130 n_jobs: -1
	Ivis	Projector	k: 10 model: szubert n_epochs_without_progress: 50 supervision_weight: 0.6616307886310997	k: 17 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7382457430509433	k: 20 model: hinton n_epochs_without_progress: 21 supervision_weight: 1.0	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 28 n_neighbors: 7 p: 4 weights: uniform	activation: tanh alpha: 0.0001 layer_one: 150 layer_two: 150 max_iter: 2000	C: 73.11319085987152 gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 50 n_jobs: -1
14	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 9 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 132 layer_two: 50 max_iter: 2000	C: 1000.0 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 86 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 30 n_neighbors: 3 p: 3 weights: distance	activation: relu alpha: 0.3384295631043929 layer_one: 127 layer_two: 146 max_iter: 2000	C: 697.8032774750423 gamma: 0.09630089417823245 kernel: rbf max_iter: 100000 probability: True	max_depth: 7 n_estimators: 50 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.9933191055835903 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.6835829624005945 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.08770743247240766 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 0.06806495353212162 kernel: linear n_jobs: -1
		Classifier	leaf_size: 19 n_neighbors: 8 p: 1 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 145 max_iter: 2000	C: 313.01955005700785 gamma: 29.566379088759547 kernel: rbf max_iter: 100000 probability: True	max_depth: 7 n_estimators: 149 n_jobs: -1
	UMAP	Projector	n_epochs: 1734 n_neighbors: 3 target_weight: 0.4770078774223391	n_epochs: 162 n_neighbors: 3 target_weight: 0.8061985596309498	n_epochs: 1071 n_neighbors: 3 target_weight: 0.6321033357464765	n_epochs: 129 n_neighbors: 3 target_weight: 0.7446428110083896
		Classifier	leaf_size: 20 n_neighbors: 28 p: 5 weights: distance	activation: tanh alpha: 0.0001 layer_one: 52 layer_two: 64 max_iter: 377	C: 1000.0 gamma: 40.933695760575844 kernel: linear max_iter: 100000 probability: True	max_depth: 13 n_estimators: 68 n_jobs: -1
	Ivis	Projector	k: 3 model: maaten n_epochs_without_progress: 50 supervision_weight: 0.6551892053061859	k: 102 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.6247758719054924	k: 30 model: hinton n_epochs_without_progress: 42 supervision_weight: 1.0
		Classifier	leaf_size: 30 n_neighbors: 3 p: 3 weights: uniform	activation: relu alpha: 9.6999856001457 layer_one: 86 layer_two: 150 max_iter: 1455	C: 353.0365785005836 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 12 n_estimators: 150 n_jobs: -1
15	PCA	Projector	—	—	—	—

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 9 p: 2 weights: distance	activation: relu alpha: 0.7600770314406469 layer_one: 150 layer_two: 91 max_iter: 2000	C: 887.1934847726662 gamma: 87.86363274508588 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 98 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 4 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 1.0263892119612559 layer_one: 61 layer_two: 150 max_iter: 2000	C: 51.93772943207991 gamma: 3.409993953690111 kernel: linear max_iter: 100000 probability: True	max_depth: 6 n_estimators: 51 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.38436309858652473 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.9922461871467445 kernel: linear n_jobs: -1
		Classifier	leaf_size: 20 n_neighbors: 3 p: 4 weights: uniform	activation: relu alpha: 0.0001 layer_one: 112 layer_two: 125 max_iter: 883	C: 1000.0 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 93 n_jobs: -1
	UMAP	Projector	n_epochs: 1002 n_neighbors: 3 target_weight: 0.36667091602972113	n_epochs: 332 n_neighbors: 3 target_weight: 0.13086695455451516	n_epochs: 1380 n_neighbors: 3 target_weight: 0.6244973407806826	n_epochs: 1694 n_neighbors: 3 target_weight: 0.0
		Classifier	leaf_size: 30 n_neighbors: 3 p: 1 weights: distance	activation: tanh alpha: 0.5456592303005244 layer_one: 108 layer_two: 50 max_iter: 1825	C: 988.2597388315379 gamma: 8.451555490492273 kernel: linear max_iter: 100000 probability: True	max_depth: 9 n_estimators: 84 n_jobs: -1
	Ivis	Projector	k: 40 model: hinton n_epochs_without_progress: 40 supervision_weight: 0.5938060833967508	k: 92 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.8416993266783988	k: 3 model: hinton n_epochs_without_progress: 44 supervision_weight: 1.0	k: 77 model: hinton n_epochs_without_progress: 42 supervision_weight: 0.7956687902848774
		Classifier	leaf_size: 3 n_neighbors: 3 p: 4 weights: distance	activation: relu alpha: 4.352006618284117 layer_one: 50 layer_two: 52 max_iter: 100	C: 523.3613469640138 gamma: 57.6532910245864 kernel: linear max_iter: 100000 probability: True	max_depth: 7 n_estimators: 59 n_jobs: -1
16	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 10 p: 3 weights: distance	activation: relu alpha: 0.9967679817983741 layer_one: 109 layer_two: 144 max_iter: 1927	C: 1000.0 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 10 n_estimators: 117 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 19 n_neighbors: 3 p: 3 weights: distance	activation: relu alpha: 0.01764259418277855 layer_one: 150 layer_two: 84 max_iter: 567	C: 172.1652277105039 gamma: 13.540014426084166 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 81 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.04928857631836522 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.44784625976518055 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.757490343993959 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 11 n_neighbors: 3 p: 4 weights: uniform	activation: relu alpha: 0.2641477240017928 layer_one: 108 layer_two: 50 max_iter: 1410	C: 582.762447275279 gamma: 1.265733895313126 kernel: rbf max_iter: 100000 probability: True	max_depth: 13 n_estimators: 64 n_jobs: -1
	UMAP	Projector	n_epochs: 1251 n_neighbors: 3 target_weight: 0.0	n_epochs: 1814 n_neighbors: 3 target_weight: 0.1487045118268598 activation: tanh	n_epochs: 1854 n_neighbors: 3 target_weight: 0.8660991665915508 C: 508.2878265355826	n_epochs: 100 n_neighbors: 3 target_weight: 0.0
		Classifier	leaf_size: 12 n_neighbors: 3 p: 5 weights: uniform	alpha: 0.0001 layer_one: 136 layer_two: 150 max_iter: 952	gamma: 80.45966664900145 kernel: linear max_iter: 100000 probability: True	max_depth: 14 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 35 supervision_weight: 0.731484808046956	k: 80 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0 activation: relu	k: 26 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.9999953560226044 C: 718.517244355067	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.6232559488560716
		Classifier	leaf_size: 8 n_neighbors: 3 p: 3 weights: uniform	alpha: 9.735129161468988 layer_one: 94 layer_two: 58 max_iter: 1083	gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
17	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 7 p: 3 weights: distance	activation: relu alpha: 0.0001 layer_one: 73 layer_two: 50 max_iter: 2000	C: 737.9129553312374 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 50 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 30 n_neighbors: 5 p: 3 weights: distance	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 50 max_iter: 1111	C: 303.619831118757 gamma: 1.383946680714243 kernel: rbf max_iter: 100000 probability: True	max_depth: 7 n_estimators: 69 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.31943090853403394 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.0 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.8190338035390964 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: cosine n_jobs: -1

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 3 n_neighbors: 3 p: 3 weights: uniform	activation: relu alpha: 0.0001 layer_one: 131 layer_two: 74 max_iter: 834	C: 890.7403742738444 gamma: 12.09713202426722 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 79 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0	n_epochs: 1407 n_neighbors: 3 target_weight: 0.7088931217745906 activation: tanh	n_epochs: 2000 n_neighbors: 3 target_weight: 0.09359593958015094 C: 1000.0	n_epochs: 2000 n_neighbors: 3 target_weight: 0.07259872310982776
		Classifier	leaf_size: 19 n_neighbors: 3 p: 4 weights: uniform	alpha: 0.0001 layer_one: 84 layer_two: 56 max_iter: 1837	gamma: 10.259090026699097 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
	Ivis	Projector	k: 39 model: hinton n_epochs_without_progress: 46 supervision_weight: 0.3927961092969755	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.8031289596058926 activation: relu	k: 69 model: hinton n_epochs_without_progress: 42 supervision_weight: 0.6810133855616355 C: 496.47155511296874	k: 48 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.5024064138758985
		Classifier	leaf_size: 7 n_neighbors: 3 p: 3 weights: uniform	alpha: 0.0001 layer_one: 150 layer_two: 50 max_iter: 1435	gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
18	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 4 n_neighbors: 9 p: 5 weights: distance	activation: relu alpha: 1.4971015598798425 layer_one: 140 layer_two: 127 max_iter: 2000	C: 316.2680615201599 gamma: 3.019170474146539 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 137 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack activation: relu	algorithm: arpack C: 673.1292453436583	algorithm: arpack
		Classifier	leaf_size: 21 n_neighbors: 5 p: 2 weights: distance	alpha: 0.8600303395369356 layer_one: 112 layer_two: 150 max_iter: 1818	gamma: 4.704561086157349 kernel: rbf max_iter: 100000 probability: True	max_depth: 12 n_estimators: 130 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.89982099110671 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.4068293072830099 kernel: sigmoid n_jobs: -1 activation: relu	eigen_solver: dense gamma: 0.9242020231910911 kernel: sigmoid n_jobs: -1 C: 1000.0	eigen_solver: dense gamma: 0.17402424552507686 kernel: cosine n_jobs: -1
		Classifier	leaf_size: 3 n_neighbors: 3 p: 3 weights: uniform	alpha: 0.0001 layer_one: 150 layer_two: 50 max_iter: 797	gamma: 0.012814401223146072 kernel: linear max_iter: 100000 probability: True	max_depth: 10 n_estimators: 119 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.3927547637413553	n_epochs: 1265 n_neighbors: 17 target_weight: 0.31047407438229746	n_epochs: 1832 n_neighbors: 3 target_weight: 0.002533949873223306	n_epochs: 2000 n_neighbors: 3 target_weight: 0.11287658045150423

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 10 n_neighbors: 3 p: 1 weights: distance	activation: relu alpha: 34.67766522192098 layer_one: 140 layer_two: 123 max_iter: 194	C: 210.32563264065965 gamma: 48.62425687074218 kernel: rbf max_iter: 100000 probability: True	max_depth: 12 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 103 model: maaten n_epochs_without_progress: 50 supervision_weight: 0.5718626091903833	k: 67 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 31 model: hinton n_epochs_without_progress: 11 supervision_weight: 0.570203172416242	k: 103 model: hinton n_epochs_without_progress: 43 supervision_weight: 0.5518689707387067
		Classifier	leaf_size: 12 n_neighbors: 3 p: 1 weights: uniform	activation: relu alpha: 9.381731528884917 layer_one: 50 layer_two: 68 max_iter: 1722	C: 1000.0 gamma: 50.09409248898745 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
19	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 9 p: 4 weights: distance	activation: relu alpha: 0.7836925976833965 layer_one: 50 layer_two: 78 max_iter: 1663	C: 618.471061209876 gamma: 28.958580536875978 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 50 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 29 n_neighbors: 9 p: 4 weights: distance	activation: relu alpha: 0.0001 layer_one: 74 layer_two: 123 max_iter: 2000	C: 749.1306669079572 gamma: 8.248903780574171 kernel: rbf max_iter: 100000 probability: True	max_depth: 10 n_estimators: 129 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.7197375107002559 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.5340893391612974 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.19684999616623836 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 0.0 kernel: linear n_jobs: -1
		Classifier	leaf_size: 21 n_neighbors: 22 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 50 max_iter: 2000	C: 491.2123555577162 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 50 n_jobs: -1
	UMAP	Projector	n_epochs: 100 n_neighbors: 3 target_weight: 0.5319760963835858	n_epochs: 100 n_neighbors: 3 target_weight: 0.47283788016230843	n_epochs: 883 n_neighbors: 3 target_weight: 0.0	n_epochs: 2000 n_neighbors: 6 target_weight: 0.0
		Classifier	leaf_size: 26 n_neighbors: 17 p: 3 weights: distance	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 150 max_iter: 100	C: 768.9135197851841 gamma: 7.182348718108412 kernel: rbf max_iter: 100000 probability: True	max_depth: 13 n_estimators: 142 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.6837164812694666	k: 33 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.5927351957768359	k: 3 model: szubert n_epochs_without_progress: 50 supervision_weight: 0.6542418636734116	k: 67 model: hinton n_epochs_without_progress: 40 supervision_weight: 0.6218511839518036

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 3 p: 1 weights: uniform	activation: tanh alpha: 3.4427496688879544 layer_one: 150 layer_two: 50 max_iter: 2000	C: 873.7107069786471 gamma: 61.00616769398791 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
20	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 9 p: 4 weights: distance	activation: relu alpha: 0.0001 layer_one: 55 layer_two: 50 max_iter: 978	C: 13.083980282220773 gamma: 2.4102653175127147 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 106 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 5 n_neighbors: 5 p: 5 weights: distance	activation: relu alpha: 1.004787589269336 layer_one: 150 layer_two: 107 max_iter: 1432	C: 72.23075419238906 gamma: 9.322742182888645 kernel: rbf max_iter: 100000 probability: True	max_depth: 7 n_estimators: 126 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.8781262357451225 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.5610827001691954 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.16809028064694756 kernel: cosine n_jobs: -1
		Classifier	leaf_size: 22 n_neighbors: 26 p: 4 weights: distance	activation: relu alpha: 0.0001 layer_one: 80 layer_two: 50 max_iter: 2000	C: 765.5286014891221 gamma: 1.5172966155870309 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 140 n_jobs: -1
	UMAP	Projector	n_epochs: 471 n_neighbors: 3 target_weight: 0.8530956396027006	n_epochs: 851 n_neighbors: 3 target_weight: 0.7060231817233346	n_epochs: 741 n_neighbors: 3 target_weight: 0.0	n_epochs: 1127 n_neighbors: 3 target_weight: 0.21130022391014386
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: uniform	activation: relu alpha: 0.0001 layer_one: 85 layer_two: 84 max_iter: 1788	C: 82.58912607892044 gamma: 64.70540304422445 kernel: rbf max_iter: 100000 probability: True	max_depth: 3 n_estimators: 75 n_jobs: -1
	Ivis	Projector	k: 28 model: hinton n_epochs_without_progress: 49 supervision_weight: 0.8438878798399373	k: 66 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 9 model: hinton n_epochs_without_progress: 46 supervision_weight: 1.0	k: 69 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7112832032726077
		Classifier	leaf_size: 19 n_neighbors: 3 p: 2 weights: uniform	activation: logistic alpha: 0.7062723279723672 layer_one: 86 layer_two: 150 max_iter: 2000	C: 863.4462621916279 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
21	PCA	Projector	—	—	—	—

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 28 n_neighbors: 9 p: 3 weights: distance	activation: relu alpha: 1.1031579035359291 layer_one: 130 layer_two: 50 max_iter: 2000	C: 108.19199565734942 gamma: 0.31583209928903416 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 89 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 29 n_neighbors: 5 p: 5 weights: distance	activation: relu alpha: 0.8152061182718124 layer_one: 51 layer_two: 81 max_iter: 2000	C: 663.8993523244598 gamma: 8.406422851016819 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 129 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.6218968763704461 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.6705200990764141 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.9063082911125077 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: cosine n_jobs: -1
		Classifier	leaf_size: 29 n_neighbors: 26 p: 5 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 150 max_iter: 340	C: 8.35339294102389 gamma: 0.7311028445919915 kernel: rbf max_iter: 100000 probability: True	max_depth: 13 n_estimators: 125 n_jobs: -1
	UMAP	Projector	n_epochs: 1393 n_neighbors: 3 target_weight: 0.6774077341900424	n_epochs: 1147 n_neighbors: 3 target_weight: 0.6956232765788818	n_epochs: 100 n_neighbors: 3 target_weight: 0.43198407699750324	n_epochs: 2000 n_neighbors: 3 target_weight: 0.1339578107177828
		Classifier	leaf_size: 3 n_neighbors: 24 p: 2 weights: distance	activation: tanh alpha: 0.12021904684730936 layer_one: 93 layer_two: 150 max_iter: 1669	C: 359.41206092155255 gamma: 10.118192900858755 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.6015463580213082	k: 33 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7436373331698098	k: 3 model: maaten n_epochs_without_progress: 50 supervision_weight: 0.6899941879925807	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.5634431808466687
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: distance	activation: relu alpha: 2.466629216838069 layer_one: 150 layer_two: 50 max_iter: 2000	C: 232.11664545811206 gamma: 30.990895904410223 kernel: linear max_iter: 100000 probability: True	max_depth: 4 n_estimators: 50 n_jobs: -1
22	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 26 n_neighbors: 5 p: 5 weights: distance	activation: relu alpha: 0.0001 layer_one: 75 layer_two: 50 max_iter: 2000	C: 544.5117438017652 gamma: 5.739770620308464 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 124 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 15 n_neighbors: 5 p: 5 weights: distance	activation: relu alpha: 0.0001 layer_one: 62 layer_two: 105 max_iter: 1901	C: 789.2971217861322 gamma: 7.7513873768002695 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 91 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.6638611310811333 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.0 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.8956230089520965 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.5998245953300442 kernel: cosine n_jobs: -1
		Classifier	leaf_size: 22 n_neighbors: 30 p: 3 weights: distance	activation: relu alpha: 1.154562374214946 layer_one: 50 layer_two: 150 max_iter: 1914	C: 743.611289882833 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 12 n_estimators: 119 n_jobs: -1
	UMAP	Projector	n_epochs: 100 n_neighbors: 3 target_weight: 0.24327614583228735	n_epochs: 100 n_neighbors: 3 target_weight: 0.49669661642651136	n_epochs: 1102 n_neighbors: 3 target_weight: 0.04067598672901601	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0
		Classifier	leaf_size: 3 n_neighbors: 24 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 85 layer_two: 118 max_iter: 1908	C: 893.0428376518295 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 7 n_estimators: 114 n_jobs: -1
	Ivis	Projector	k: 59 model: szubert n_epochs_without_progress: 50 supervision_weight: 0.6515385240645006	k: 3 model: hinton n_epochs_without_progress: 31 supervision_weight: 1.0	k: 103 model: hinton n_epochs_without_progress: 10 supervision_weight: 1.0	k: 20 model: hinton n_epochs_without_progress: 35 supervision_weight: 0.16636082758010112
		Classifier	leaf_size: 30 n_neighbors: 3 p: 2 weights: uniform	activation: tanh alpha: 7.664426104229446 layer_one: 82 layer_two: 150 max_iter: 2000	C: 693.4750120858436 gamma: 41.98769217324516 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 87 n_jobs: -1
23	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 6 n_neighbors: 5 p: 5 weights: distance	activation: relu alpha: 0.0001 layer_one: 135 layer_two: 50 max_iter: 392	C: 904.2503479040583 gamma: 8.475018936686695 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 67 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 8 n_neighbors: 5 p: 5 weights: distance	activation: relu alpha: 1.3504627516770324 layer_one: 55 layer_two: 67 max_iter: 2000	C: 721.270886595827 gamma: 5.109596656014504 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 137 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.6247904637951849 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.45432875834915176 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.6876030667339553 kernel: sigmoid n_jobs: -1

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 3 n_neighbors: 30 p: 3 weights: distance	activation: relu alpha: 0.0001 layer_one: 106 layer_two: 50 max_iter: 1863	C: 793.8406575311825 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 9 n_estimators: 100 n_jobs: -1
	UMAP	Projector	n_epochs: 547 n_neighbors: 3 target_weight: 0.7568683685855977	n_epochs: 1266 n_neighbors: 3 target_weight: 0.0	n_epochs: 743 n_neighbors: 3 target_weight: 0.3827342681674129	n_epochs: 100 n_neighbors: 3 target_weight: 0.5619590658512514
		Classifier	leaf_size: 30 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 79 layer_two: 143 max_iter: 497	C: 74.12586052690078 gamma: 32.51561266778865 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 78 n_jobs: -1
	Ivis	Projector	k: 60 model: hinton n_epochs_without_progress: 45 supervision_weight: 0.6220792665344778	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.85125268570438	k: 53 model: hinton n_epochs_without_progress: 31 supervision_weight: 1.0	k: 7 model: hinton n_epochs_without_progress: 31 supervision_weight: 0.5813999273302493
		Classifier	leaf_size: 25 n_neighbors: 3 p: 1 weights: uniform	activation: relu alpha: 8.222953389434982 layer_one: 50 layer_two: 150 max_iter: 1210	C: 231.10268219268485 gamma: 38.318136884497434 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 91 n_jobs: -1
24	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 11 n_neighbors: 5 p: 5 weights: distance	activation: relu alpha: 0.0001 layer_one: 112 layer_two: 150 max_iter: 1145	C: 751.352922297756 gamma: 0.1230359570596693 kernel: rbf max_iter: 100000 probability: True	max_depth: 12 n_estimators: 78 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 3 n_neighbors: 3 p: 5 weights: distance	activation: relu alpha: 1.360392160562942 layer_one: 150 layer_two: 124 max_iter: 1813	C: 607.9549818342551 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 150 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.4708116812112437 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.6400747537145748 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.8639085939310037 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 29 n_neighbors: 30 p: 5 weights: distance	activation: tanh alpha: 0.0001 layer_one: 50 layer_two: 50 max_iter: 2000	C: 1000.0 gamma: 71.4898405940304 kernel: linear max_iter: 100000 probability: True	max_depth: 6 n_estimators: 50 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0	n_epochs: 1021 n_neighbors: 3 target_weight: 0.0	n_epochs: 833 n_neighbors: 3 target_weight: 0.10404739425823398	n_epochs: 1318 n_neighbors: 3 target_weight: 0.6134740887411696

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 27 n_neighbors: 3 p: 5 weights: uniform	activation: tanh alpha: 0.0001 layer_one: 68 layer_two: 82 max_iter: 1380	C: 217.0646734734775 gamma: 16.210766886323416 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 22 model: hinton n_epochs_without_progress: 34 supervision_weight: 0.7113926500132782	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 103 model: hinton n_epochs_without_progress: 32 supervision_weight: 0.7872396870514169
		Classifier	leaf_size: 15 n_neighbors: 4 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 75 layer_two: 50 max_iter: 100	C: 555.3702964417649 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 3 n_estimators: 150 n_jobs: -1
25	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 5 p: 5 weights: distance	activation: relu alpha: 1.6814717478567272 layer_one: 144 layer_two: 110 max_iter: 1035	C: 945.3399535347639 gamma: 4.288137197110706 kernel: rbf max_iter: 100000 probability: True	max_depth: 12 n_estimators: 63 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 3 n_neighbors: 3 p: 4 weights: distance	activation: relu alpha: 1.4494500498420 62 layer_one: 150 layer_two: 132 max_iter: 2000	C: 523.3106464973985 gamma: 5.199600092624396 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 133 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.48688844058955233 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.7921560706871633 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.3327146701395875 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.07742676435890701 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 28 n_neighbors: 3 p: 3 weights: distance	activation: relu alpha: 5.375172208666404 layer_one: 144 layer_two: 105 max_iter: 1741	C: 155.92491284802608 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 8 n_estimators: 116 n_jobs: -1
	UMAP	Projector	n_epochs: 100 n_neighbors: 3 target_weight: 0.7200789259038205	n_epochs: 355 n_neighbors: 3 target_weight: 0.50113880141351	n_epochs: 2000 n_neighbors: 3 target_weight: 0.12691397532292892	n_epochs: 2000 n_neighbors: 3 target_weight: 0.303731093702877
		Classifier	leaf_size: 3 n_neighbors: 3 p: 5 weights: uniform	activation: relu alpha: 0.0001 layer_one: 138 layer_two: 50 max_iter: 144	C: 997.2108574002331 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 3 n_estimators: 103 n_jobs: -1
	Ivis	Projector	k: 3 model: maaten n_epochs_without_progress: 48 supervision_weight: 0.7022575998664887	k: 55 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 27 model: hinton n_epochs_without_progress: 42 supervision_weight: 1.0

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 3 n_neighbors: 12 p: 4 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 50 max_iter: 395	C: 1000.0 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 50 n_jobs: -1
26	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 4 n_neighbors: 5 p: 4 weights: distance	activation: tanh alpha: 0.0001 layer_one: 50 layer_two: 50 max_iter: 1311	C: 365.3117222648627 gamma: 8.234202246377716 kernel: rbf max_iter: 100000 probability: True	max_depth: 7 n_estimators: 150 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 4 n_neighbors: 3 p: 4 weights: distance	activation: tanh alpha: 0.0001 layer_one: 50 layer_two: 84 max_iter: 1729	C: 1000.0 gamma: 5.573027108218652 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 70 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.4451994598731434 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.6928137657456935 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.6257851609399248 kernel: rbf n_jobs: -1
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: distance	activation: tanh alpha: 0.0001 layer_one: 70 layer_two: 62 max_iter: 705	C: 774.7815570915276 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 4 n_estimators: 136 n_jobs: -1
	UMAP	Projector	n_epochs: 1602 n_neighbors: 3 target_weight: 0.0	n_epochs: 1727 n_neighbors: 3 target_weight: 0.20101062722864316	n_epochs: 429 n_neighbors: 3 target_weight: 0.14410081637415376	n_epochs: 2000 n_neighbors: 3 target_weight: 0.13316944471915698
		Classifier	leaf_size: 30 n_neighbors: 3 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 117 layer_two: 115 max_iter: 2000	C: 444.2981316884088 gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
	Ivis	Projector	k: 58 model: hinton n_epochs_without_progress: 29 supervision_weight: 0.611111211226016	k: 66 model: maaten n_epochs_without_progress: 50 supervision_weight: 1.0	k: 103 model: hinton n_epochs_without_progress: 24 supervision_weight: 1.0	k: 3 model: maaten n_epochs_without_progress: 33 supervision_weight: 0.5581230755798268
		Classifier	leaf_size: 30 n_neighbors: 3 p: 2 weights: distance	activation: tanh alpha: 0.0001 layer_one: 91 layer_two: 82 max_iter: 858	C: 824.4987389372027 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 134 n_jobs: -1
27	PCA	Projector	—	—	—	—

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 29 n_neighbors: 9 p: 3 weights: distance	activation: relu alpha: 1.6039818205780954 layer_one: 138 layer_two: 57 max_iter: 1062	C: 707.7983045554797 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 78 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 29 n_neighbors: 4 p: 3 weights: uniform	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 144 max_iter: 1868	C: 893.6807512566943 gamma: 8.44256081898936 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 72 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 1.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.5133614459012131 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.6780975713228571 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: uniform	activation: tanh alpha: 1.9245490574813313 layer_one: 118 layer_two: 102 max_iter: 220	C: 142.83566663042444 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0	n_epochs: 1403 n_neighbors: 3 target_weight: 0.4021171399399041	n_epochs: 157 n_neighbors: 3 target_weight: 0.0	n_epochs: 2000 n_neighbors: 3 target_weight: 0.22526977021068154
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: distance	activation: logistic alpha: 0.0001 layer_one: 96 layer_two: 88 max_iter: 1984	C: 976.8411228492345 gamma: 77.06707545609012 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 75 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 28 supervision_weight: 0.47348790107079214	k: 3 model: maaten n_epochs_without_progress: 50 supervision_weight: 0.6726093813194556	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 91 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7023072505586685
		Classifier	leaf_size: 30 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 150 max_iter: 100	C: 559.903612797365 gamma: 72.23409881388885 kernel: linear max_iter: 100000 probability: True	max_depth: 9 n_estimators: 112 n_jobs: -1
28	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 9 p: 3 weights: distance	activation: relu alpha: 1.5011935808749324 layer_one: 150 layer_two: 50 max_iter: 382	C: 11.047033831963073 gamma: 62.282039235955345 kernel: linear max_iter: 100000 probability: True	max_depth: 10 n_estimators: 72 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 5 n_neighbors: 4 p: 5 weights: distance	activation: relu alpha: 1.4816155145153964 layer_one: 139 layer_two: 135 max_iter: 1170	C: 645.8791409644442 gamma: 1.432018165539536 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 121 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.9258464710780319 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.45729033474398784 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 0.17277926363320092 kernel: linear n_jobs: -1
		Classifier	leaf_size: 28 n_neighbors: 3 p: 5 weights: uniform	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 150 max_iter: 100	C: 1000.0 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 10 n_estimators: 74 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.2647274139830837	n_epochs: 2000 n_neighbors: 3 target_weight: 0.4628463599670748	n_epochs: 1193 n_neighbors: 3 target_weight: 0.44573840123170133	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0
		Classifier	leaf_size: 3 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 111 layer_two: 123 max_iter: 1850	C: 1000.0 gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 3 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7245128379165925	k: 65 model: hinton n_epochs_without_progress: 27 supervision_weight: 1.0	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7441637949721523	k: 86 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.4021586205488932
		Classifier	leaf_size: 30 n_neighbors: 3 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 90 layer_two: 76 max_iter: 100	C: 436.8659167130019 gamma: 92.47783199189122 kernel: linear max_iter: 100000 probability: True	max_depth: 10 n_estimators: 50 n_jobs: -1
29	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 9 p: 4 weights: distance	activation: relu alpha: 0.9582626122466034 layer_one: 140 layer_two: 126 max_iter: 805	C: 4.840517649502485 gamma: 0.35475815192711935 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 150 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 30 n_neighbors: 5 p: 2 weights: uniform	activation: relu alpha: 0.0001 layer_one: 140 layer_two: 50 max_iter: 2000	C: 153.65719434425034 gamma: 11.272110126232901 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 150 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.18037461189886328 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.06030385449830394 kernel: sigmoid n_jobs: -1

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 27 n_neighbors: 17 p: 2 weights: distance	activation: tanh alpha: 0.0001 layer_one: 50 layer_two: 150 max_iter: 100	C: 790.0962895527706 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 60 n_jobs: -1
	UMAP	Projector	n_epochs: 100 n_neighbors: 3 target_weight: 0.49734669798551995	n_epochs: 1306 n_neighbors: 20 target_weight: 0.16608759300929726 activation: relu	n_epochs: 1824 n_neighbors: 3 target_weight: 0.6687355956985024 C: 285.8526561325675	n_epochs: 100 n_neighbors: 3 target_weight: 0.20641590076561284
		Classifier	leaf_size: 3 n_neighbors: 30 p: 1 weights: distance	alpha: 29.25439784560575 layer_one: 101 layer_two: 60 max_iter: 100	gamma: 80.29218014141348 kernel: linear max_iter: 100000 probability: True	max_depth: 3 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 38 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.705287131608921	k: 59 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0 activation: relu	k: 3 model: szubert n_epochs_without_progress: 50 supervision_weight: 0.7779013568509313 C: 88.96009725580322	k: 75 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.638172103623781
		Classifier	leaf_size: 8 n_neighbors: 3 p: 1 weights: uniform	alpha: 0.0001 layer_one: 150 layer_two: 69 max_iter: 388	gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 14 n_estimators: 150 n_jobs: -1
30	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 9 p: 5 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 72 max_iter: 100	C: 8.528026023383928 gamma: 0.798889892895192 kernel: rbf max_iter: 100000 probability: True	max_depth: 14 n_estimators: 124 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack activation: relu	algorithm: arpack C: 655.3829312682367	algorithm: arpack
		Classifier	leaf_size: 3 n_neighbors: 3 p: 5 weights: distance	alpha: 1.2267265075320792 layer_one: 50 layer_two: 69 max_iter: 1665	gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 110 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.0920272392173876 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.2882331131583511 kernel: cosine n_jobs: -1 activation: relu	eigen_solver: dense gamma: 0.3993317605374972 kernel: sigmoid n_jobs: -1 C: 575.029257655428	eigen_solver: dense gamma: 0.06236249387693095 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 27 n_neighbors: 8 p: 5 weights: uniform	alpha: 2.963265635319119 layer_one: 85 layer_two: 108 max_iter: 1555	gamma: 8.171654821919143 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 125 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0	n_epochs: 310 n_neighbors: 3 target_weight: 0.6010887958542774	n_epochs: 688 n_neighbors: 13 target_weight: 0.34289802528068225	n_epochs: 771 n_neighbors: 3 target_weight: 0.0

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 97 layer_two: 89 max_iter: 1381	C: 535.0469604378168 gamma: 72.14118582103833 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 56 n_jobs: -1
	Ivis	Projector	k: 43 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7719106702021828	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.6642330356457625	k: 101 model: szubert n_epochs_without_progress: 41 supervision_weight: 0.6969958514800445	k: 22 model: hinton n_epochs_without_progress: 29 supervision_weight: 0.399770558669278
		Classifier	leaf_size: 30 n_neighbors: 3 p: 1 weights: uniform	activation: relu alpha: 2.4091539954349797 layer_one: 67 layer_two: 150 max_iter: 2000	C: 287.49527698842184 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 150 n_jobs: -1
31	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 9 p: 4 weights: distance	activation: relu alpha: 1.7634586091383033 layer_one: 150 layer_two: 96 max_iter: 2000	C: 20.609553378808123 gamma: 0.8044266096186804 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 140 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 29 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 71 max_iter: 723	C: 379.83673993189 gamma: 0.17439838718262735 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 50 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.3974925199563221 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.5439849617974107 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.02157296101352624 kernel: linear n_jobs: -1
		Classifier	leaf_size: 26 n_neighbors: 12 p: 2 weights: distance	activation: relu alpha: 3.5782675584593227 layer_one: 100 layer_two: 76 max_iter: 2000	C: 527.7145303946247 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 4 n_estimators: 97 n_jobs: -1
	UMAP	Projector	n_epochs: 100 n_neighbors: 3 target_weight: 0.22099538866695592	n_epochs: 2000 n_neighbors: 3 target_weight: 0.479447672957177	n_epochs: 100 n_neighbors: 3 target_weight: 0.33048144245131506	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0
		Classifier	leaf_size: 3 n_neighbors: 3 p: 5 weights: distance	activation: relu alpha: 0.0001 layer_one: 144 layer_two: 150 max_iter: 1148	C: 652.0797561363885 gamma: 43.44769222924176 kernel: rbf max_iter: 100000 probability: True	max_depth: 12 n_estimators: 68 n_jobs: -1
	Ivis	Projector	k: 25 model: hinton n_epochs_without_progress: 32 supervision_weight: 0.6959502487024996	k: 3 model: hinton n_epochs_without_progress: 34 supervision_weight: 0.5061377253462203	k: 25 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 11 model: hinton n_epochs_without_progress: 41 supervision_weight: 0.5314607432499362

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: distance	activation: tanh alpha: 0.0001 layer_one: 50 layer_two: 50 max_iter: 2000	C: 916.9635106820444 gamma: 39.21098746398091 kernel: linear max_iter: 100000 probability: True	max_depth: 14 n_estimators: 102 n_jobs: -1
32	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 10 p: 5 weights: distance	activation: tanh alpha: 1.676174365649308 layer_one: 122 layer_two: 150 max_iter: 100	C: 669.4874875095779 gamma: 0.04459652820361258 kernel: rbf max_iter: 100000 probability: True	max_depth: 4 n_estimators: 50 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 29 n_neighbors: 7 p: 5 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 59 max_iter: 991	C: 66.40537952000012 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 9 n_estimators: 131 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.17663134463792277 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.4054750631562665 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.515248023784009 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.7854511941009966 kernel: linear n_jobs: -1
		Classifier	leaf_size: 30 n_neighbors: 15 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 150 max_iter: 2000	C: 97.4204356142179 gamma: 73.51690723764952 kernel: linear max_iter: 100000 probability: True	max_depth: 4 n_estimators: 50 n_jobs: -1
	UMAP	Projector	n_epochs: 100 n_neighbors: 3 target_weight: 0.37827219379027843	n_epochs: 2000 n_neighbors: 3 target_weight: 0.2541317660272477	n_epochs: 2000 n_neighbors: 3 target_weight: 0.30827534308545135	n_epochs: 2000 n_neighbors: 3 target_weight: 0.16638385255631893
		Classifier	leaf_size: 3 n_neighbors: 30 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 150 layer_two: 60 max_iter: 1405	C: 881.7522601871086 gamma: 0.017902851077226542 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 43 supervision_weight: 1.0	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 100 model: hinton n_epochs_without_progress: 42 supervision_weight: 0.7795913717597893
		Classifier	leaf_size: 30 n_neighbors: 15 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 57 layer_two: 50 max_iter: 1055	C: 683.3213229868969 gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 3 n_estimators: 119 n_jobs: -1
33	PCA	Projector	—	—	—	—

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 29 n_neighbors: 9 p: 3 weights: distance	activation: relu alpha: 2.727852876815387 layer_one: 50 layer_two: 150 max_iter: 240	C: 20.083052380084307 gamma: 0.728635399630508 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 150 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 3 n_neighbors: 3 p: 3 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 150 max_iter: 795	C: 252.98003557863484 gamma: 40.100951069395514 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 63 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.8905722119357365 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.9436144798844686 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.43954265855306024 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.2903073435135491 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 12 n_neighbors: 17 p: 3 weights: distance	activation: relu alpha: 1.589572009125536 layer_one: 50 layer_two: 150 max_iter: 1955	C: 270.97574129534695 gamma: 45.8409146521341 kernel: linear max_iter: 100000 probability: True	max_depth: 4 n_estimators: 150 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.02390412848015136	n_epochs: 1043 n_neighbors: 3 target_weight: 0.0 activation: tanh	n_epochs: 100 n_neighbors: 3 target_weight: 0.3579302892398693 C: 595.6599383012966	n_epochs: 100 n_neighbors: 3 target_weight: 0.15506089303488643
		Classifier	leaf_size: 28 n_neighbors: 3 p: 2 weights: uniform	alpha: 0.0001 layer_one: 127 layer_two: 150 max_iter: 587	gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 13 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 35 supervision_weight: 0.6812326566884336	k: 49 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.5278222410921197	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0
		Classifier	leaf_size: 4 n_neighbors: 3 p: 1 weights: uniform	activation: relu alpha: 3.2409171812990527 layer_one: 150 layer_two: 150 max_iter: 2000	C: 1000.0 gamma: 79.32190462254586 kernel: linear max_iter: 100000 probability: True	max_depth: 12 n_estimators: 50 n_jobs: -1
34	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 9 p: 4 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 132 max_iter: 396	C: 333.9875491880478 gamma: 24.597229616128022 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 148 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 12 n_neighbors: 9 p: 3 weights: distance	activation: relu alpha: 0.0001 layer_one: 122 layer_two: 50 max_iter: 550	C: 182.8302762600897 gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 4 n_estimators: 115 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.649139362858185 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.7567933843018042 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.24559374406357284 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 3 n_neighbors: 15 p: 5 weights: distance	activation: relu alpha: 3.1336321232428808 layer_one: 150 layer_two: 118 max_iter: 314	C: 828.9876069393897 gamma: 79.61256013130485 kernel: linear max_iter: 100000 probability: True	max_depth: 12 n_estimators: 67 n_jobs: -1
	UMAP	Projector	n_epochs: 1774 n_neighbors: 3 target_weight: 0.0	n_epochs: 741 n_neighbors: 25 target_weight: 0.0	n_epochs: 369 n_neighbors: 3 target_weight: 0.634226984381916	n_epochs: 1532 n_neighbors: 3 target_weight: 0.24915976777833926
		Classifier	leaf_size: 15 n_neighbors: 3 p: 5 weights: distance	activation: relu alpha: 0.0001 layer_one: 108 layer_two: 65 max_iter: 1008	C: 462.8721062956498 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 10 n_estimators: 67 n_jobs: -1
	Ivis	Projector	k: 82 model: hinton n_epochs_without_progress: 37 supervision_weight: 0.995931531694064	k: 94 model: hinton n_epochs_without_progress: 49 supervision_weight: 0.6992025089646564	k: 3 model: hinton n_epochs_without_progress: 10 supervision_weight: 1.0	k: 28 model: hinton n_epochs_without_progress: 41 supervision_weight: 0.29274464167180453
		Classifier	leaf_size: 19 n_neighbors: 3 p: 3 weights: distance	activation: relu alpha: 5.442565888125904 layer_one: 150 layer_two: 130 max_iter: 1963	C: 915.2300626472287 gamma: 25.210590198851214 kernel: linear max_iter: 100000 probability: True	max_depth: 14 n_estimators: 87 n_jobs: -1
35	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 26 n_neighbors: 6 p: 2 weights: uniform	activation: relu alpha: 0.0001 layer_one: 102 layer_two: 50 max_iter: 503	C: 435.6297807753962 gamma: 0.09346427732189042 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 108 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 29 n_neighbors: 9 p: 5 weights: distance	activation: relu alpha: 0.5407483325969135 layer_one: 72 layer_two: 50 max_iter: 1265	C: 6.851591363291902 gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 7 n_estimators: 111 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.31099840806405776 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.7804098328461436 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.5896309953737657 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.3341748057991762 kernel: linear n_jobs: -1

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: distance	activation: tanh alpha: 6.6038734168009166 layer_one: 86 layer_two: 100 max_iter: 541	C: 288.8035616287915 gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 9 n_estimators: 105 n_jobs: -1
	UMAP	Projector	n_epochs: 1944 n_neighbors: 3 target_weight: 0.0	n_epochs: 2000 n_neighbors: 3 target_weight: 0.15694048762157778	n_epochs: 576 n_neighbors: 3 target_weight: 0.0	n_epochs: 510 n_neighbors: 3 target_weight: 0.8645191754720486
		Classifier	leaf_size: 12 n_neighbors: 4 p: 5 weights: uniform	activation: tanh alpha: 0.0001 layer_one: 126 layer_two: 150 max_iter: 885	C: 162.52584333480294 gamma: 44.686090542895016 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 87 n_jobs: -1
	Ivis	Projector	k: 46 model: hinton n_epochs_without_progress: 42 supervision_weight: 0.572138093388645	k: 58 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 56 model: hinton n_epochs_without_progress: 29 supervision_weight: 1.0	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0
		Classifier	leaf_size: 27 n_neighbors: 3 p: 1 weights: uniform	activation: tanh alpha: 0.8594933193270056 layer_one: 147 layer_two: 50 max_iter: 2000	C: 1000.0 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 3 n_estimators: 150 n_jobs: -1
36	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 26 n_neighbors: 3 p: 4 weights: distance	activation: relu alpha: 2.2186782168610786 layer_one: 66 layer_two: 149 max_iter: 1963	C: 689.3899342048948 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 10 n_estimators: 116 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 15 n_neighbors: 9 p: 5 weights: distance	activation: relu alpha: 1.293939035016844 layer_one: 57 layer_two: 64 max_iter: 1128	C: 578.6377217589866 gamma: 47.852515285207105 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 80 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.7775747210419661 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.8044843414596874 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.5053513950473708 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.7206290793268647 kernel: rbf n_jobs: -1
		Classifier	leaf_size: 27 n_neighbors: 26 p: 2 weights: distance	activation: relu alpha: 2.123537453855658 layer_one: 127 layer_two: 59 max_iter: 560	C: 129.38817745703483 gamma: 10.881738668933309 kernel: linear max_iter: 100000 probability: True	max_depth: 4 n_estimators: 88 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.11621986400013387	n_epochs: 1509 n_neighbors: 17 target_weight: 0.16004502648834631	n_epochs: 100 n_neighbors: 3 target_weight: 0.6902650036154895	n_epochs: 1945 n_neighbors: 3 target_weight: 0.0

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: distance	activation: logistic alpha: 0.0001 layer_one: 54 layer_two: 108 max_iter: 1382	C: 496.0244727955185 gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
	Ivis	Projector	k: 3 model: szubert n_epochs_without_progress: 50 supervision_weight: 0.5883433972232973	k: 19 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 103 model: hinton n_epochs_without_progress: 40 supervision_weight: 0.8482001631868566	k: 103 model: hinton n_epochs_without_progress: 47 supervision_weight: 0.7001022139110502
		Classifier	leaf_size: 18 n_neighbors: 3 p: 4 weights: distance	activation: relu alpha: 13.06949107799844 layer_one: 150 layer_two: 137 max_iter: 280	C: 535.679069721333 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 98 n_jobs: -1
37	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 11 n_neighbors: 3 p: 3 weights: distance	activation: relu alpha: 0.0001 layer_one: 78 layer_two: 72 max_iter: 100	C: 79.69760462768588 gamma: 93.33278282644804 kernel: linear max_iter: 100000 probability: True	max_depth: 11 n_estimators: 92 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 7 n_neighbors: 3 p: 5 weights: distance	activation: relu alpha: 1.2869049173206124 layer_one: 50 layer_two: 99 max_iter: 750	C: 159.59756384855655 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 11 n_estimators: 50 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.1854436970311414 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.9671973190941219 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.21368883689862686 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 27 n_neighbors: 13 p: 5 weights: distance	activation: relu alpha: 2.5631022054198804 layer_one: 71 layer_two: 107 max_iter: 2000	C: 267.3879290439756 gamma: 14.520593965102346 kernel: linear max_iter: 100000 probability: True	max_depth: 7 n_estimators: 50 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.4293894247911524	n_epochs: 1637 n_neighbors: 3 target_weight: 0.5009825023056397	n_epochs: 1951 n_neighbors: 3 target_weight: 0.8659726020191841	n_epochs: 1725 n_neighbors: 3 target_weight: 0.0
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: uniform	activation: tanh alpha: 0.0001 layer_one: 150 layer_two: 150 max_iter: 2000	C: 15.967551506513203 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 7 n_estimators: 150 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 37 supervision_weight: 0.7332569171287163	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.8801144523719147	k: 54 model: hinton n_epochs_without_progress: 44 supervision_weight: 0.7013279259140393	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7404890832079203

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 3 n_neighbors: 3 p: 5 weights: uniform	activation: tanh alpha: 0.0001 layer_one: 76 layer_two: 50 max_iter: 1541	C: 163.40016060044053 gamma: 3.039399504070872 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
38	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 5 p: 4 weights: distance	activation: relu alpha: 1.6116828519723625 layer_one: 140 layer_two: 119 max_iter: 1737	C: 69.42799089033623 gamma: 54.725288258452274 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 135 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 4 n_neighbors: 4 p: 3 weights: distance	activation: logistic alpha: 0.0001 layer_one: 144 layer_two: 150 max_iter: 1605	C: 862.888178513386 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 13 n_estimators: 96 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.304000111434382 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.9864681932955266 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.03771536767978271 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.7203305512755042 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 3 n_neighbors: 3 p: 2 weights: distance	activation: tanh alpha: 4.638893069001675 layer_one: 144 layer_two: 150 max_iter: 414	C: 68.56938788772527 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
	UMAP	Projector	n_epochs: 1901 n_neighbors: 3 target_weight: 0.9382366426176266	n_epochs: 2000 n_neighbors: 3 target_weight: 0.8193297947013539	n_epochs: 351 n_neighbors: 3 target_weight: 0.6133253167357574	n_epochs: 100 n_neighbors: 3 target_weight: 0.0
		Classifier	leaf_size: 14 n_neighbors: 30 p: 4 weights: distance	activation: tanh alpha: 0.0001 layer_one: 92 layer_two: 124 max_iter: 100	C: 128.41311247396925 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 86 n_jobs: -1
	Ivis	Projector	k: 95 model: hinton n_epochs_without_progress: 38 supervision_weight: 0.7944479652254066	k: 61 model: hinton n_epochs_without_progress: 29 supervision_weight: 0.8906478623966728	k: 103 model: hinton n_epochs_without_progress: 29 supervision_weight: 1.0	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.5830011069886086
		Classifier	leaf_size: 19 n_neighbors: 5 p: 5 weights: uniform	activation: tanh alpha: 6.285853410236815 layer_one: 105 layer_two: 139 max_iter: 2000	C: 306.8545008466727 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 141 n_jobs: -1
39	PCA	Projector	—	—	—	—

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 29 n_neighbors: 3 p: 2 weights: distance	activation: tanh alpha: 1.2123595319782965 layer_one: 57 layer_two: 150 max_iter: 1064	C: 10.368625650531026 gamma: 32.13708258553747 kernel: linear max_iter: 100000 probability: True	max_depth: 6 n_estimators: 145 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 3 n_neighbors: 12 p: 2 weights: distance	activation: relu alpha: 1.765128343407209 layer_one: 50 layer_two: 150 max_iter: 1488	C: 318.3768618231486 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 56 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.9439002310847002 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.24262799151715525 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.45855396067267096 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.1319740984883452 kernel: rbf n_jobs: -1
		Classifier	leaf_size: 26 n_neighbors: 3 p: 5 weights: distance	activation: tanh alpha: 6.548230978047802 layer_one: 78 layer_two: 118 max_iter: 1116	C: 898.1492073478906 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 7 n_estimators: 91 n_jobs: -1
	UMAP	Projector	n_epochs: 880 n_neighbors: 3 target_weight: 0.0	n_epochs: 1276 n_neighbors: 3 target_weight: 0.28094172475928475	n_epochs: 1602 n_neighbors: 3 target_weight: 0.545078432397389	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: uniform	activation: logistic alpha: 0.0001 layer_one: 127 layer_two: 129 max_iter: 100	C: 1000.0 gamma: 3.6915340310080906 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.5190831877910796	k: 79 model: hinton n_epochs_without_progress: 48 supervision_weight: 0.5510251239198957	k: 36 model: maaten n_epochs_without_progress: 43 supervision_weight: 0.337756616275813	k: 3 model: hinton n_epochs_without_progress: 37 supervision_weight: 0.7563316943883025
		Classifier	leaf_size: 3 n_neighbors: 3 p: 1 weights: distance	activation: tanh alpha: 5.7173659822906995 layer_one: 50 layer_two: 118 max_iter: 1102	C: 113.00986314101196 gamma: 90.2136571887322 kernel: linear max_iter: 100000 probability: True	max_depth: 4 n_estimators: 150 n_jobs: -1
40	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 3 p: 5 weights: distance	activation: relu alpha: 1.9324828525802207 layer_one: 50 layer_two: 144 max_iter: 521	C: 667.074782546718 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 101 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 3 n_neighbors: 9 p: 2 weights: distance	activation: logistic alpha: 0.0001 layer_one: 150 layer_two: 80 max_iter: 2000	C: 850.2273917309577 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 10 n_estimators: 68 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.09353298040685297 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.6790782952031977 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 18 n_neighbors: 14 p: 2 weights: distance	activation: relu alpha: 3.382638215019798 layer_one: 50 layer_two: 150 max_iter: 2000	C: 379.0542839101242 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
	UMAP	Projector	n_epochs: 1797 n_neighbors: 3 target_weight: 0.12838951316911662	n_epochs: 100 n_neighbors: 3 target_weight: 0.9718538330903349	n_epochs: 1631 n_neighbors: 3 target_weight: 0.3704450579977015	n_epochs: 2000 n_neighbors: 3 target_weight: 0.35303373225808443
		Classifier	leaf_size: 30 n_neighbors: 3 p: 1 weights: distance	activation: tanh alpha: 2.5923536200778536 layer_one: 55 layer_two: 83 max_iter: 595	C: 1000.0 gamma: 38.22751656406525 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.6917656993451302	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.9796768248410346	k: 3 model: hinton n_epochs_without_progress: 24 supervision_weight: 1.0	k: 36 model: hinton n_epochs_without_progress: 40 supervision_weight: 0.538565730473347
		Classifier	leaf_size: 15 n_neighbors: 3 p: 1 weights: uniform	activation: relu alpha: 34.71985560010954 layer_one: 66 layer_two: 150 max_iter: 100	C: 711.7393470720953 gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 96 n_jobs: -1
41	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 28 n_neighbors: 3 p: 3 weights: distance	activation: relu alpha: 0.0001 layer_one: 81 layer_two: 150 max_iter: 1947	C: 1000.0 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 116 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 28 n_neighbors: 7 p: 2 weights: distance	activation: logistic alpha: 0.0001 layer_one: 120 layer_two: 50 max_iter: 2000	C: 335.63994274467336 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 12 n_estimators: 123 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.5132662241016347 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.49056290829053345 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.9770790345693425 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.01614464666682031 kernel: linear n_jobs: -1

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 25 n_neighbors: 3 p: 2 weights: distance	activation: tanh alpha: 3.2458227625978044 layer_one: 50 layer_two: 50 max_iter: 2000	C: 114.0874921038062 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 144 n_jobs: -1
	UMAP	Projector	n_epochs: 1559 n_neighbors: 3 target_weight: 0.190499098808805	n_epochs: 100 n_neighbors: 3 target_weight: 0.0 activation: tanh	n_epochs: 2000 n_neighbors: 3 target_weight: 0.4193396244342387 C: 1000.0	n_epochs: 1916 n_neighbors: 3 target_weight: 0.3525071606010566
		Classifier	leaf_size: 30 n_neighbors: 6 p: 5 weights: distance	alpha: 0.0001 layer_one: 150 layer_two: 116 max_iter: 2000	gamma: 44.20534067719033 kernel: linear max_iter: 100000 probability: True	max_depth: 11 n_estimators: 148 n_jobs: -1
	Ivis	Projector	k: 23 model: hinton n_epochs_without_progress: 34 supervision_weight: 0.774492375782169	k: 77 model: hinton n_epochs_without_progress: 35 supervision_weight: 0.6221906038993933 activation: relu	k: 3 model: szubert n_epochs_without_progress: 50 supervision_weight: 0.522053623926726 C: 613.3709500419756	k: 6 model: maaten n_epochs_without_progress: 50 supervision_weight: 0.430497241980297
		Classifier	leaf_size: 26 n_neighbors: 3 p: 5 weights: distance	alpha: 0.0001 layer_one: 150 layer_two: 148 max_iter: 964	gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 4 n_estimators: 150 n_jobs: -1
42	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 4 p: 4 weights: distance	activation: relu alpha: 3.648546017583118 layer_one: 50 layer_two: 121 max_iter: 1986	C: 1000.0 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 142 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 4 n_neighbors: 4 p: 2 weights: distance	activation: relu alpha: 0.7293187490294178 layer_one: 50 layer_two: 61 max_iter: 1027	C: 782.9024682467793 gamma: 4.045288985718835 kernel: linear max_iter: 100000 probability: True	max_depth: 5 n_estimators: 124 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.006753666842547312 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.25389589785515426 kernel: sigmoid n_jobs: -1	eigen_solver: dense gamma: 0.3351054225513454 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 0.15081039716149092 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 30 n_neighbors: 3 p: 2 weights: distance	activation: tanh alpha: 7.848262507327729 layer_one: 57 layer_two: 146 max_iter: 2000	C: 41.47598428484377 gamma: 95.92188813363893 kernel: linear max_iter: 100000 probability: True	max_depth: 7 n_estimators: 150 n_jobs: -1
	UMAP	Projector	n_epochs: 100 n_neighbors: 3 target_weight: 0.0	n_epochs: 867 n_neighbors: 3 target_weight: 0.7446642543861863	n_epochs: 122 n_neighbors: 3 target_weight: 0.7980352089656328	n_epochs: 100 n_neighbors: 3 target_weight: 0.6998470053137006

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 3 p: 2 weights: distance	activation: tanh alpha: 10.414479751845294 layer_one: 67 layer_two: 50 max_iter: 1227	C: 388.660203442472 gamma: 92.230838351338 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 52 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.5051006586905933	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.7666184826665834	k: 3 model: maaten n_epochs_without_progress: 50 supervision_weight: 0.5810331034818683	k: 103 model: hinton n_epochs_without_progress: 48 supervision_weight: 0.5083441800442722
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: distance	activation: tanh alpha: 1.4021509100634468 layer_one: 52 layer_two: 50 max_iter: 924	C: 346.5656481867306 gamma: 71.30569425698084 kernel: linear max_iter: 100000 probability: True	max_depth: 12 n_estimators: 82 n_jobs: -1
43	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 30 n_neighbors: 3 p: 4 weights: distance	activation: relu alpha: 4.219774255598379 layer_one: 66 layer_two: 118 max_iter: 1258	C: 906.6286726708863 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 7 n_estimators: 141 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 30 n_neighbors: 4 p: 2 weights: distance	activation: relu alpha: 3.5896765594341 layer_one: 137 layer_two: 146 max_iter: 1555	C: 254.21932270660605 gamma: 0.01 kernel: sigmoid max_iter: 100000 probability: True	max_depth: 12 n_estimators: 98 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.8596401479823259 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.801147243820887 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.05131306529520384 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 0.20042800786836817 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 28 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 91 layer_two: 73 max_iter: 1262	C: 52.44189523301812 gamma: 92.92295915882114 kernel: linear max_iter: 100000 probability: True	max_depth: 13 n_estimators: 122 n_jobs: -1
	UMAP	Projector	n_epochs: 1810 n_neighbors: 3 target_weight: 0.0	n_epochs: 278 n_neighbors: 3 target_weight: 0.6863652171805222	n_epochs: 100 n_neighbors: 3 target_weight: 0.836153584710153	n_epochs: 156 n_neighbors: 3 target_weight: 0.24963530518808705
		Classifier	leaf_size: 15 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 135 max_iter: 1595	C: 1000.0 gamma: 100.0 kernel: linear max_iter: 100000 probability: True	max_depth: 12 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 73 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.5710388679641806	k: 103 model: hinton n_epochs_without_progress: 40 supervision_weight: 0.8050220916237644	k: 103 model: szubert n_epochs_without_progress: 36 supervision_weight: 1.0	k: 51 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 3 n_neighbors: 3 p: 3 weights: uniform	activation: tanh alpha: 0.0001 layer_one: 150 layer_two: 67 max_iter: 2000	C: 351.4969802411639 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 13 n_estimators: 125 n_jobs: -1
44	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 5 n_neighbors: 6 p: 3 weights: distance	activation: relu alpha: 0.5166405079028297 layer_one: 112 layer_two: 118 max_iter: 1147	C: 1000.0 gamma: 0.10066641558497286 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 118 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 30 n_neighbors: 7 p: 2 weights: distance	activation: relu alpha: 0.16031974614926622 layer_one: 109 layer_two: 127 max_iter: 100	C: 304.9716857738144 gamma: 0.01 kernel: sigmoid max_iter: 100000 probability: True	max_depth: 9 n_estimators: 140 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.9236508918360096 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.1560854959335263 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.04777325576168397 kernel: linear n_jobs: -1
		Classifier	leaf_size: 23 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 3.6857572106847387 layer_one: 139 layer_two: 65 max_iter: 182	C: 167.83407307967963 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 56 n_jobs: -1
	UMAP	Projector	n_epochs: 1867 n_neighbors: 3 target_weight: 0.015738611003770296	n_epochs: 1755 n_neighbors: 3 target_weight: 0.6173811668526161	n_epochs: 1440 n_neighbors: 3 target_weight: 0.5087526321778221	n_epochs: 1630 n_neighbors: 3 target_weight: 0.0760695360600486
		Classifier	leaf_size: 30 n_neighbors: 3 p: 1 weights: distance	activation: tanh alpha: 0.0001 layer_one: 104 layer_two: 85 max_iter: 1266	C: 647.8965818407091 gamma: 10.852766742676781 kernel: linear max_iter: 100000 probability: True	max_depth: 9 n_estimators: 109 n_jobs: -1
	Ivis	Projector	k: 51 model: hinton n_epochs_without_progress: 36 supervision_weight: 0.9966655400520242	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 18 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.219205345703382
		Classifier	leaf_size: 30 n_neighbors: 3 p: 3 weights: uniform	activation: relu alpha: 0.0001 layer_one: 75 layer_two: 90 max_iter: 2000	C: 783.0188262932629 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 10 n_estimators: 64 n_jobs: -1
45	PCA	Projector	—	—	—	—

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 6 p: 3 weights: distance	activation: relu alpha: 3.4175456327381526 layer_one: 52 layer_two: 72 max_iter: 1739	C: 816.0503944241881 gamma: 0.01 kernel: sigmoid max_iter: 100000 probability: True	max_depth: 11 n_estimators: 138 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 3 n_neighbors: 7 p: 2 weights: distance	activation: tanh alpha: 0.0001 layer_one: 50 layer_two: 150 max_iter: 2000	C: 3.34669467241798 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 4 n_estimators: 55 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.37210033403821613 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.16353566286220475 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 0.4026119244986983 kernel: rbf n_jobs: -1	eigen_solver: dense gamma: 0.034858778196293135 kernel: rbf n_jobs: -1
		Classifier	leaf_size: 23 n_neighbors: 4 p: 3 weights: distance	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 141 max_iter: 1960	C: 986.6427266778213 gamma: 39.440107454149434 kernel: linear max_iter: 100000 probability: True	max_depth: 11 n_estimators: 62 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.6554203766150037	n_epochs: 1765 n_neighbors: 3 target_weight: 0.8849661585582255	n_epochs: 1790 n_neighbors: 3 target_weight: 0.2985456324007674	n_epochs: 936 n_neighbors: 3 target_weight: 0.4123332314383286
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: distance	activation: logistic alpha: 0.0001 layer_one: 95 layer_two: 113 max_iter: 1581	C: 895.6292824164341 gamma: 97.0856312093697 kernel: linear max_iter: 100000 probability: True	max_depth: 3 n_estimators: 50 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.786344784848558	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.6137820031979043	k: 88 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 69 model: hinton n_epochs_without_progress: 37 supervision_weight: 0.8786633670890368
		Classifier	leaf_size: 3 n_neighbors: 3 p: 2 weights: uniform	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 92 max_iter: 529	C: 821.2929119339933 gamma: 5.405660610074156 kernel: rbf max_iter: 100000 probability: True	max_depth: 15 n_estimators: 50 n_jobs: -1
46	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 0.682626030778527 layer_one: 50 layer_two: 61 max_iter: 2000	C: 97.71874144488723 gamma: 57.12753600872881 kernel: linear max_iter: 100000 probability: True	max_depth: 12 n_estimators: 77 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 11 n_neighbors: 7 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 50 layer_two: 150 max_iter: 100	C: 816.4693734336851 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 127 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.08257826851974912 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.8999388096445281 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: linear n_jobs: -1
		Classifier	leaf_size: 24 n_neighbors: 12 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 70 layer_two: 133 max_iter: 615	C: 574.8954637136158 gamma: 0.01 kernel: sigmoid max_iter: 100000 probability: True	max_depth: 15 n_estimators: 76 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0	n_epochs: 583 n_neighbors: 3 target_weight: 0.08453303724556759	n_epochs: 2000 n_neighbors: 3 target_weight: 0.14247388713764988	n_epochs: 2000 n_neighbors: 3 target_weight: 0.22856087717939447
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: uniform	activation: relu alpha: 0.0001 layer_one: 106 layer_two: 51 max_iter: 677	C: 907.45641856912 gamma: 64.53550018798933 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.3956242790894832	k: 3 model: maaten n_epochs_without_progress: 50 supervision_weight: 0.6318600439782127	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 43 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.4273688828972146
		Classifier	leaf_size: 30 n_neighbors: 3 p: 1 weights: distance	activation: relu alpha: 0.0001 layer_one: 59 layer_two: 144 max_iter: 1108	C: 757.8635953463709 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 15 n_estimators: 150 n_jobs: -1
47	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 4 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 0.22855207947590797 layer_one: 149 layer_two: 97 max_iter: 1890	C: 99.29147134413904 gamma: 0.21711008057910355 kernel: rbf max_iter: 100000 probability: True	max_depth: 6 n_estimators: 150 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 3 n_neighbors: 7 p: 2 weights: distance	activation: relu alpha: 3.8235840777794476 layer_one: 148 layer_two: 87 max_iter: 1206	C: 677.9993188938923 gamma: 0.01 kernel: sigmoid max_iter: 100000 probability: True	max_depth: 9 n_estimators: 79 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.30889263882481427 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.303763432157457 kernel: linear n_jobs: -1	eigen_solver: dense gamma: 0.4891445711688449 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.52445803384717 kernel: linear n_jobs: -1

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 10 p: 2 weights: distance	activation: relu alpha: 0.0001 layer_one: 83 layer_two: 50 max_iter: 1151	C: 9.370708993099655 gamma: 4.52036994659513 kernel: linear max_iter: 100000 probability: True	max_depth: 10 n_estimators: 140 n_jobs: -1
	UMAP	Projector	n_epochs: 100 n_neighbors: 3 target_weight: 0.7952639573604607	n_epochs: 2000 n_neighbors: 18 target_weight: 0.0	n_epochs: 358 n_neighbors: 3 target_weight: 0.35161232995100555	n_epochs: 1818 n_neighbors: 3 target_weight: 0.004016080365088502
		Classifier	leaf_size: 18 n_neighbors: 3 p: 3 weights: distance	activation: tanh alpha: 0.0001 layer_one: 88 layer_two: 50 max_iter: 663	C: 601.3017590118619 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 11 n_estimators: 96 n_jobs: -1
	Ivis	Projector	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 16 model: hinton n_epochs_without_progress: 48 supervision_weight: 0.5736441415125082	k: 62 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0	k: 10 model: szubert n_epochs_without_progress: 31 supervision_weight: 0.5090715066731741
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: uniform	activation: tanh alpha: 9.930909464602754 layer_one: 150 layer_two: 150 max_iter: 2000	C: 301.74366705857665 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 8 n_estimators: 141 n_jobs: -1
48	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 5 p: 4 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 120 max_iter: 100	C: 392.9212566963307 gamma: 0.01 kernel: sigmoid max_iter: 100000 probability: True	max_depth: 11 n_estimators: 150 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack	algorithm: arpack	algorithm: arpack
		Classifier	leaf_size: 30 n_neighbors: 7 p: 2 weights: distance	activation: relu alpha: 0.052803714325660236 layer_one: 114 layer_two: 122 max_iter: 339	C: 153.2572046746775 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 9 n_estimators: 85 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.07435999782984802 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.17522599438129643 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 0.21692216709387385 kernel: sigmoid n_jobs: -1
		Classifier	leaf_size: 7 n_neighbors: 3 p: 2 weights: distance	activation: relu alpha: 3.6746522121706793 layer_one: 150 layer_two: 104 max_iter: 147	C: 32.83911752132349 gamma: 0.01 kernel: linear max_iter: 100000 probability: True	max_depth: 3 n_estimators: 150 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0	n_epochs: 126 n_neighbors: 3 target_weight: 0.6226924605506605	n_epochs: 614 n_neighbors: 3 target_weight: 0.8582585840656572	n_epochs: 1065 n_neighbors: 3 target_weight: 0.03816048232986867

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 26 n_neighbors: 3 p: 1 weights: distance	activation: tanh alpha: 0.42662337785527776 layer_one: 121 layer_two: 125 max_iter: 761	C: 322.6855543490742 gamma: 77.18554762423796 kernel: rbf max_iter: 100000 probability: True	max_depth: 3 n_estimators: 103 n_jobs: -1
	Ivis	Projector	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.6866066778256388	k: 100 model: hinton n_epochs_without_progress: 50 supervision_weight: 1.0 activation: relu	k: 26 model: hinton n_epochs_without_progress: 48 supervision_weight: 0.8403391337430599 C: 1000.0	k: 69 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.9958051656529847
		Classifier	leaf_size: 30 n_neighbors: 3 p: 5 weights: distance	alpha: 3.5920435190067117 layer_one: 87 layer_two: 135 max_iter: 100	gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 5 n_estimators: 133 n_jobs: -1
49	PCA	Projector	—	—	—	—
		Classifier	leaf_size: 3 n_neighbors: 5 p: 4 weights: distance	activation: relu alpha: 0.0001 layer_one: 150 layer_two: 132 max_iter: 100	C: 4.2459166484753394 gamma: 94.13843081865907 kernel: linear max_iter: 100000 probability: True	max_depth: 4 n_estimators: 64 n_jobs: -1
	TSVD	Projector	algorithm: arpack	algorithm: arpack activation: tanh	algorithm: arpack C: 329.59360036341343	algorithm: arpack
		Classifier	leaf_size: 30 n_neighbors: 7 p: 1 weights: distance	alpha: 3.92558793380252 layer_one: 106 layer_two: 135 max_iter: 2000	gamma: 0.01 kernel: sigmoid max_iter: 100000 probability: True	max_depth: 8 n_estimators: 147 n_jobs: -1
	KPCA	Projector	eigen_solver: dense gamma: 0.36085783007895245 kernel: cosine n_jobs: -1	eigen_solver: dense gamma: 1.0 kernel: cosine n_jobs: -1 activation: relu	eigen_solver: dense gamma: 0.8230884060355343 kernel: cosine n_jobs: -1 C: 736.9256361849311	eigen_solver: dense gamma: 0.058146842639825516 kernel: linear n_jobs: -1
		Classifier	leaf_size: 11 n_neighbors: 3 p: 2 weights: distance	alpha: 0.23445735614125152 layer_one: 150 layer_two: 82 max_iter: 1569	gamma: 0.01 kernel: sigmoid max_iter: 100000 probability: True	max_depth: 5 n_estimators: 126 n_jobs: -1
	UMAP	Projector	n_epochs: 2000 n_neighbors: 3 target_weight: 0.0	n_epochs: 1685 n_neighbors: 3 target_weight: 0.785234894491651 activation: relu	n_epochs: 1609 n_neighbors: 3 target_weight: 0.13323707675796437 C: 886.2366948114285	n_epochs: 100 n_neighbors: 3 target_weight: 0.688273833857079
		Classifier	leaf_size: 3 n_neighbors: 3 p: 5 weights: uniform	alpha: 0.0001 layer_one: 97 layer_two: 115 max_iter: 1653	gamma: 91.79539182452324 kernel: linear max_iter: 100000 probability: True	max_depth: 9 n_estimators: 51 n_jobs: -1
	Ivis	Projector	k: 41 model: hinton n_epochs_without_progress: 28 supervision_weight: 0.4312278125223603	k: 3 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.6686165475675948	k: 59 model: hinton n_epochs_without_progress: 10 supervision_weight: 0.21237602432806701	k: 103 model: hinton n_epochs_without_progress: 50 supervision_weight: 0.5417026143518382

Continued on next page

Table 4 – Hyper-parameters of best-performing pipelines obtained by means of Bayesian optimization with 10-fold stratified cross-validation.

		Classifier	KNN	MLP	SVM	RF
Dimensionality	Projector	Parameter				
		Classifier	leaf_size: 30 n_neighbors: 3 p: 4 weights: uniform	activation: tanh alpha: 4.288980243510306 layer_one: 147 layer_two: 53 max_iter: 1417	C: 1000.0 gamma: 0.01 kernel: rbf max_iter: 100000 probability: True	max_depth: 3 n_estimators: 150 n_jobs: -1

APPENDIX

D

INDEPENDENT TEST MEASUREMENTS

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
50	—	KNN	0.706522	0.727273	0.706522	0.417911	0.782609	0.630435
		MLP	0.717391	0.750000	0.717391	0.450377	0.847826	0.586957
		SVM	0.793478	0.804124	0.793478	0.590455	0.847826	0.739130
		RF	0.804348	0.820000	0.804348	0.618115	0.891304	0.717391
2	PCA	KNN	0.706522	0.747664	0.706522	0.436926	0.869565	0.543478
		MLP	0.717391	0.754717	0.717391	0.456435	0.869565	0.565217
		SVM	0.706522	0.756757	0.706522	0.453539	0.913043	0.500000
		RF	0.728261	0.761905	0.728261	0.475923	0.869565	0.586957
	TSVD	KNN	0.739130	0.764706	0.739130	0.489979	0.847826	0.630435
		MLP	0.663043	0.699029	0.663043	0.335830	0.782609	0.543478
		SVM	0.673913	0.722222	0.673913	0.370991	0.847826	0.500000
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
	KPCA	KNN	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		MLP	0.684783	0.707071	0.684783	0.373920	0.760870	0.608696
		SVM	0.706522	0.756757	0.706522	0.453539	0.913043	0.500000
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
	UMAP	KNN	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
		MLP	0.717391	0.763636	0.717391	0.472456	0.913043	0.521739
		SVM	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
		RF	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
	Ivis	KNN	0.717391	0.750000	0.717391	0.450377	0.847826	0.586957
		MLP	0.804348	0.812500	0.804348	0.611010	0.847826	0.760870
		SVM	0.739130	0.764706	0.739130	0.489979	0.847826	0.630435
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
3	PCA	KNN	0.706522	0.752294	0.706522	0.444513	0.891304	0.521739
		MLP	0.717391	0.767857	0.717391	0.482805	0.934783	0.500000
		SVM	0.684783	0.743363	0.684783	0.415376	0.913043	0.456522
		RF	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
	TSVD	KNN	0.673913	0.727273	0.673913	0.377964	0.869565	0.478261
		MLP	0.663043	0.715596	0.663043	0.350931	0.847826	0.478261

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		SVM	0.673913	0.732143	0.673913	0.386244	0.891304	0.456522
		RF	0.684783	0.712871	0.684783	0.376848	0.782609	0.586957
	KPCA	KNN	0.695652	0.735849	0.695652	0.410792	0.847826	0.543478
		MLP	0.695652	0.730769	0.695652	0.405340	0.826087	0.565217
		SVM	0.706522	0.737864	0.706522	0.425385	0.826087	0.586957
		RF	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
	UMAP	KNN	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		RF	0.663043	0.693069	0.663043	0.332513	0.760870	0.565217
	Ivis	KNN	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696
		MLP	0.684783	0.728972	0.684783	0.390934	0.847826	0.521739
		SVM	0.695652	0.725490	0.695652	0.400892	0.804348	0.586957
		RF	0.760870	0.784314	0.760870	0.534522	0.869565	0.652174
4	PCA	KNN	0.728261	0.766355	0.728261	0.482918	0.891304	0.565217
		MLP	0.717391	0.763636	0.717391	0.472456	0.913043	0.521739
		SVM	0.706522	0.737864	0.706522	0.425385	0.826087	0.586957
		RF	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
	TSVD	KNN	0.673913	0.722222	0.673913	0.370991	0.847826	0.500000
		MLP	0.510870	0.117647	0.510870	0.047946	0.065217	0.956522
		SVM	0.728261	0.770642	0.728261	0.491304	0.913043	0.543478
		RF	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696
	KPCA	KNN	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.717391	0.750000	0.717391	0.450377	0.847826	0.586957
		RF	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
	UMAP	KNN	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		MLP	0.695652	0.725490	0.695652	0.400892	0.804348	0.586957
		SVM	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		RF	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
	Ivis	KNN	0.728261	0.770642	0.728261	0.491304	0.913043	0.543478
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.782609	0.803922	0.782609	0.579066	0.891304	0.673913
		RF	0.684783	0.707071	0.684783	0.373920	0.760870	0.608696
5	PCA	KNN	0.684783	0.718447	0.684783	0.380608	0.804348	0.565217
		MLP	0.663043	0.686869	0.663043	0.329929	0.739130	0.586957
		SVM	0.717391	0.763636	0.717391	0.472456	0.913043	0.521739
		RF	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
	TSVD	KNN	0.684783	0.728972	0.684783	0.390934	0.847826	0.521739
		MLP	0.684783	0.747826	0.684783	0.426737	0.934783	0.434783
		SVM	0.706522	0.761062	0.706522	0.464244	0.934783	0.478261
		RF	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
	KPCA	KNN	0.684783	0.718447	0.684783	0.380608	0.804348	0.565217
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.717391	0.763636	0.717391	0.472456	0.913043	0.521739
		RF	0.695652	0.708333	0.695652	0.392792	0.739130	0.652174
	UMAP	KNN	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
		RF	0.739130	0.744681	0.739130	0.478714	0.760870	0.717391
	Ivis	KNN	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
6	PCA	KNN	0.750000	0.792793	0.750000	0.549021	0.956522	0.543478
		MLP	0.684783	0.728972	0.684783	0.390934	0.847826	0.521739
		SVM	0.739130	0.773585	0.739130	0.502079	0.891304	0.586957
		RF	0.684783	0.707071	0.684783	0.373920	0.760870	0.608696
	TSVD	KNN	0.717391	0.750000	0.717391	0.450377	0.847826	0.586957
		MLP	0.673913	0.705882	0.673913	0.356348	0.782609	0.565217
		SVM	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
		RF	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
	KPCA	KNN	0.717391	0.754717	0.717391	0.456435	0.869565	0.565217
		MLP	0.673913	0.680851	0.673913	0.348155	0.695652	0.652174
		SVM	0.728261	0.770642	0.728261	0.491304	0.913043	0.543478
		RF	0.684783	0.701031	0.684783	0.371768	0.739130	0.630435
	UMAP	KNN	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
	Ivis	KNN	0.728261	0.761905	0.728261	0.475923	0.869565	0.586957
		MLP	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696
		SVM	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		RF	0.673913	0.693878	0.673913	0.350823	0.739130	0.608696
7	PCA	KNN	0.673913	0.722222	0.673913	0.370991	0.847826	0.500000
		MLP	0.684783	0.728972	0.684783	0.390934	0.847826	0.521739
		SVM	0.706522	0.737864	0.706522	0.425385	0.826087	0.586957
		RF	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
	TSVD	KNN	0.706522	0.752294	0.706522	0.444513	0.891304	0.521739
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.717391	0.759259	0.717391	0.463739	0.891304	0.543478
		RF	0.706522	0.727273	0.706522	0.417911	0.782609	0.630435
	KPCA	KNN	0.663043	0.686869	0.663043	0.329929	0.739130	0.586957
		MLP	0.706522	0.696629	0.706522	0.413925	0.673913	0.739130
		SVM	0.597826	0.633663	0.597826	0.199508	0.695652	0.500000
		RF	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
	UMAP	KNN	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
		MLP	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
		SVM	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		RF	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
	Ivis	KNN	0.760870	0.788462	0.760870	0.540453	0.891304	0.630435
		MLP	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
		SVM	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
		RF	0.695652	0.725490	0.695652	0.400892	0.804348	0.586957

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
8	PCA	KNN	0.717391	0.759259	0.717391	0.463739	0.891304	0.543478
		MLP	0.684783	0.728972	0.684783	0.390934	0.847826	0.521739
		SVM	0.684783	0.718447	0.684783	0.380608	0.804348	0.565217
		RF	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
	TSVD	KNN	0.663043	0.680412	0.663043	0.328031	0.717391	0.608696
		MLP	0.706522	0.715789	0.706522	0.413925	0.739130	0.673913
		SVM	0.717391	0.750000	0.717391	0.450377	0.847826	0.586957
		RF	0.706522	0.715789	0.706522	0.413925	0.739130	0.673913
	KPCA	KNN	0.706522	0.715789	0.706522	0.413925	0.739130	0.673913
		MLP	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696
		SVM	0.760870	0.788462	0.760870	0.540453	0.891304	0.630435
		RF	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
	UMAP	KNN	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		MLP	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
		SVM	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		RF	0.728261	0.731183	0.728261	0.456630	0.739130	0.717391
	Ivis	KNN	0.739130	0.764706	0.739130	0.489979	0.847826	0.630435
		MLP	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
		SVM	0.706522	0.715789	0.706522	0.413925	0.739130	0.673913
		RF	0.793478	0.808081	0.793478	0.593873	0.869565	0.717391
9	PCA	KNN	0.695652	0.745455	0.695652	0.425210	0.891304	0.500000
		MLP	0.652174	0.741935	0.652174	0.423659	1.000000	0.304348
		SVM	0.663043	0.693069	0.663043	0.332513	0.760870	0.565217
		RF	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
	TSVD	KNN	0.652174	0.673469	0.652174	0.306970	0.717391	0.586957
		MLP	0.717391	0.759259	0.717391	0.463739	0.891304	0.543478
		SVM	0.717391	0.759259	0.717391	0.463739	0.891304	0.543478
		RF	0.739130	0.773585	0.739130	0.502079	0.891304	0.586957
	KPCA	KNN	0.641304	0.666667	0.641304	0.285939	0.717391	0.565217
		MLP	0.684783	0.718447	0.684783	0.380608	0.804348	0.565217
		SVM	0.706522	0.727273	0.706522	0.417911	0.782609	0.630435
		RF	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
	UMAP	KNN	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		MLP	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		SVM	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
	Ivis	KNN	0.695652	0.730769	0.695652	0.405340	0.826087	0.565217
		MLP	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
		SVM	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		RF	0.793478	0.800000	0.793478	0.588209	0.826087	0.760870
10	PCA	KNN	0.619565	0.660194	0.619565	0.246276	0.739130	0.500000
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
		RF	0.706522	0.727273	0.706522	0.417911	0.782609	0.630435
	TSVD	KNN	0.684783	0.688172	0.684783	0.369653	0.695652	0.673913
		MLP	0.684783	0.728972	0.684783	0.390934	0.847826	0.521739
		SVM	0.706522	0.742857	0.706522	0.430597	0.847826	0.565217

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		RF	0.760870	0.784314	0.760870	0.534522	0.869565	0.652174
	KPCA	KNN	0.597826	0.593407	0.597826	0.195698	0.586957	0.608696
		MLP	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
		SVM	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		RF	0.706522	0.721649	0.706522	0.415505	0.760870	0.652174
	UMAP	KNN	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
		MLP	0.728261	0.725275	0.728261	0.456630	0.717391	0.739130
		SVM	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		RF	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
	Ivis	KNN	0.695652	0.730769	0.695652	0.405340	0.826087	0.565217
		MLP	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
		SVM	0.728261	0.761905	0.728261	0.475923	0.869565	0.586957
		RF	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
11	PCA	KNN	0.663043	0.699029	0.663043	0.335830	0.782609	0.543478
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.706522	0.737864	0.706522	0.425385	0.826087	0.586957
		RF	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
	TSVD	KNN	0.684783	0.707071	0.684783	0.373920	0.760870	0.608696
		MLP	0.695652	0.708333	0.695652	0.392792	0.739130	0.652174
		SVM	0.630435	0.630435	0.630435	0.260870	0.630435	0.630435
		RF	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
	KPCA	KNN	0.695652	0.708333	0.695652	0.392792	0.739130	0.652174
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.565217	0.555556	0.565217	0.130558	0.543478	0.586957
		RF	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
	UMAP	KNN	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
		MLP	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
		SVM	0.771739	0.778947	0.771739	0.544638	0.804348	0.739130
		RF	0.717391	0.723404	0.717391	0.435194	0.739130	0.695652
	Ivis	KNN	0.739130	0.773585	0.739130	0.502079	0.891304	0.586957
		MLP	0.760870	0.784314	0.760870	0.534522	0.869565	0.652174
		SVM	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
12	PCA	KNN	0.652174	0.686275	0.652174	0.311805	0.760870	0.543478
		MLP	0.478261	0.000000	0.478261	-0.149071	0.000000	0.956522
		SVM	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696
		RF	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
	TSVD	KNN	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
		MLP	0.673913	0.716981	0.673913	0.365148	0.826087	0.521739
		SVM	0.695652	0.730769	0.695652	0.405340	0.826087	0.565217
		RF	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
	KPCA	KNN	0.652174	0.686275	0.652174	0.311805	0.760870	0.543478
		MLP	0.663043	0.747967	0.663043	0.441367	1.000000	0.326087
		SVM	0.652174	0.666667	0.652174	0.305505	0.695652	0.608696
		RF	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
	UMAP	KNN	0.760870	0.788462	0.760870	0.540453	0.891304	0.630435
		MLP	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		SVM	0.793478	0.815534	0.793478	0.604494	0.913043	0.673913
		RF	0.706522	0.721649	0.706522	0.415505	0.760870	0.652174
	Ivis	KNN	0.673913	0.711538	0.673913	0.360302	0.804348	0.543478
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		RF	0.771739	0.783505	0.771739	0.546718	0.826087	0.717391
13	PCA	KNN	0.652174	0.692308	0.652174	0.315264	0.782609	0.521739
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.750000	0.741573	0.750000	0.501067	0.717391	0.782609
		RF	0.684783	0.681319	0.684783	0.369653	0.673913	0.695652
	TSVD	KNN	0.652174	0.686275	0.652174	0.311805	0.760870	0.543478
		MLP	0.673913	0.605263	0.673913	0.370991	0.500000	0.847826
		SVM	0.706522	0.703297	0.706522	0.413141	0.695652	0.717391
		RF	0.684783	0.701031	0.684783	0.371768	0.739130	0.630435
	KPCA	KNN	0.695652	0.725490	0.695652	0.400892	0.804348	0.586957
		MLP	0.641304	0.731707	0.641304	0.382518	0.978261	0.304348
		SVM	0.663043	0.673684	0.663043	0.326783	0.695652	0.630435
		RF	0.695652	0.688889	0.695652	0.391675	0.673913	0.717391
	UMAP	KNN	0.760870	0.784314	0.760870	0.534522	0.869565	0.652174
		MLP	0.728261	0.712644	0.728261	0.459243	0.673913	0.782609
		SVM	0.673913	0.693878	0.673913	0.350823	0.739130	0.608696
		RF	0.695652	0.720000	0.695652	0.397360	0.782609	0.608696
	Ivis	KNN	0.739130	0.764706	0.739130	0.489979	0.847826	0.630435
		MLP	0.695652	0.702128	0.695652	0.391675	0.717391	0.673913
		SVM	0.771739	0.783505	0.771739	0.546718	0.826087	0.717391
		RF	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
14	PCA	KNN	0.652174	0.692308	0.652174	0.315264	0.782609	0.521739
		MLP	0.652174	0.680000	0.652174	0.309058	0.739130	0.565217
		SVM	0.641304	0.679612	0.641304	0.291053	0.760870	0.521739
		RF	0.684783	0.694737	0.684783	0.370354	0.717391	0.652174
	TSVD	KNN	0.673913	0.700000	0.673913	0.353209	0.760870	0.586957
		MLP	0.663043	0.699029	0.663043	0.335830	0.782609	0.543478
		SVM	0.695652	0.735849	0.695652	0.410792	0.847826	0.543478
		RF	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
	KPCA	KNN	0.663043	0.680412	0.663043	0.328031	0.717391	0.608696
		MLP	0.684783	0.688172	0.684783	0.369653	0.695652	0.673913
		SVM	0.728261	0.719101	0.728261	0.457496	0.695652	0.760870
		RF	0.695652	0.695652	0.695652	0.391304	0.695652	0.695652
	UMAP	KNN	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		MLP	0.663043	0.626506	0.663043	0.332513	0.565217	0.760870
		SVM	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
		RF	0.750000	0.752688	0.750000	0.500118	0.760870	0.739130
	Ivis	KNN	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
15	PCA	KNN	0.652174	0.692308	0.652174	0.315264	0.782609	0.521739

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		MLP	0.706522	0.715789	0.706522	0.413925	0.739130	0.673913
		SVM	0.630435	0.666667	0.630435	0.267261	0.739130	0.521739
		RF	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
	TSVD	KNN	0.695652	0.725490	0.695652	0.400892	0.804348	0.586957
		MLP	0.510870	0.671533	0.510870	0.104828	1.000000	0.021739
		SVM	0.663043	0.704762	0.663043	0.339945	0.804348	0.521739
		RF	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
	KPCA	KNN	0.641304	0.666667	0.641304	0.285939	0.717391	0.565217
		MLP	0.750000	0.752688	0.750000	0.500118	0.760870	0.739130
		SVM	0.652174	0.698113	0.652174	0.319505	0.804348	0.500000
		RF	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
	UMAP	KNN	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
		MLP	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		SVM	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		RF	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
	Ivis	KNN	0.739130	0.769231	0.739130	0.495415	0.869565	0.608696
		MLP	0.673913	0.745763	0.673913	0.421637	0.956522	0.391304
		SVM	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
		RF	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
16	PCA	KNN	0.652174	0.703704	0.652174	0.324617	0.826087	0.478261
		MLP	0.641304	0.736000	0.641304	0.405656	1.000000	0.282609
		SVM	0.641304	0.673267	0.641304	0.288178	0.739130	0.543478
		RF	0.739130	0.744681	0.739130	0.478714	0.760870	0.717391
	TSVD	KNN	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696
		MLP	0.641304	0.691589	0.641304	0.298949	0.804348	0.478261
		SVM	0.728261	0.719101	0.728261	0.457496	0.695652	0.760870
		RF	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
	KPCA	KNN	0.663043	0.686869	0.663043	0.329929	0.739130	0.586957
		MLP	0.706522	0.721649	0.706522	0.415505	0.760870	0.652174
		SVM	0.717391	0.717391	0.717391	0.434783	0.717391	0.717391
		RF	0.728261	0.719101	0.728261	0.457496	0.695652	0.760870
	UMAP	KNN	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696
		MLP	0.706522	0.727273	0.706522	0.417911	0.782609	0.630435
		SVM	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		RF	0.663043	0.699029	0.663043	0.335830	0.782609	0.543478
	Ivis	KNN	0.739130	0.769231	0.739130	0.495415	0.869565	0.608696
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		RF	0.771739	0.783505	0.771739	0.546718	0.826087	0.717391
17	PCA	KNN	0.608696	0.653846	0.608696	0.225189	0.739130	0.478261
		MLP	0.652174	0.659574	0.652174	0.304636	0.673913	0.630435
		SVM	0.641304	0.679612	0.641304	0.291053	0.760870	0.521739
		RF	0.695652	0.681818	0.695652	0.392792	0.652174	0.739130
	TSVD	KNN	0.630435	0.673077	0.630435	0.270226	0.760870	0.500000
		MLP	0.717391	0.717391	0.717391	0.434783	0.717391	0.717391
		SVM	0.695652	0.702128	0.695652	0.391675	0.717391	0.673913
		RF	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
	KPCA	KNN	0.663043	0.686869	0.663043	0.329929	0.739130	0.586957
		MLP	0.684783	0.688172	0.684783	0.369653	0.695652	0.673913
		SVM	0.652174	0.680000	0.652174	0.309058	0.739130	0.565217
		RF	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
	UMAP	KNN	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
		MLP	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		SVM	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
	Ivis	KNN	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
		MLP	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
		SVM	0.684783	0.701031	0.684783	0.371768	0.739130	0.630435
		RF	0.782609	0.787234	0.782609	0.565752	0.804348	0.760870
18	PCA	KNN	0.619565	0.672897	0.619565	0.252957	0.782609	0.456522
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.706522	0.703297	0.706522	0.413141	0.695652	0.717391
		RF	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
	TSVD	KNN	0.652174	0.698113	0.652174	0.319505	0.804348	0.500000
		MLP	0.521739	0.153846	0.521739	0.088045	0.086957	0.956522
		SVM	0.706522	0.715789	0.706522	0.413925	0.739130	0.673913
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
	KPCA	KNN	0.673913	0.693878	0.673913	0.350823	0.739130	0.608696
		MLP	0.706522	0.682353	0.706522	0.417911	0.630435	0.782609
		SVM	0.673913	0.673913	0.673913	0.347826	0.673913	0.673913
		RF	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
	UMAP	KNN	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		RF	0.760870	0.765957	0.760870	0.522233	0.782609	0.739130
	Ivis	KNN	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
		RF	0.793478	0.808081	0.793478	0.593873	0.869565	0.717391
19	PCA	KNN	0.652174	0.698113	0.652174	0.319505	0.804348	0.500000
		MLP	0.641304	0.736000	0.641304	0.405656	1.000000	0.282609
		SVM	0.717391	0.697674	0.717391	0.438529	0.652174	0.782609
		RF	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
	TSVD	KNN	0.641304	0.697248	0.641304	0.304140	0.826087	0.456522
		MLP	0.706522	0.747664	0.706522	0.436926	0.869565	0.543478
		SVM	0.695652	0.695652	0.695652	0.391304	0.695652	0.695652
		RF	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
	KPCA	KNN	0.706522	0.737864	0.706522	0.425385	0.826087	0.586957
		MLP	0.728261	0.761905	0.728261	0.475923	0.869565	0.586957
		SVM	0.673913	0.700000	0.673913	0.353209	0.760870	0.586957
		RF	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
	UMAP	KNN	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		MLP	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		SVM	0.695652	0.720000	0.695652	0.397360	0.782609	0.608696

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		RF	0.630435	0.653061	0.630435	0.263117	0.695652	0.565217
	Ivis	KNN	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		RF	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
20	PCA	KNN	0.652174	0.698113	0.652174	0.319505	0.804348	0.500000
		MLP	0.684783	0.694737	0.684783	0.370354	0.717391	0.652174
		SVM	0.739130	0.769231	0.739130	0.495415	0.869565	0.608696
		RF	0.750000	0.752688	0.750000	0.500118	0.760870	0.739130
	TSVD	KNN	0.673913	0.716981	0.673913	0.365148	0.826087	0.521739
		MLP	0.663043	0.743802	0.663043	0.420084	0.978261	0.347826
		SVM	0.706522	0.703297	0.706522	0.413141	0.695652	0.717391
		RF	0.739130	0.764706	0.739130	0.489979	0.847826	0.630435
	KPCA	KNN	0.663043	0.710280	0.663043	0.344942	0.826087	0.500000
		MLP	0.717391	0.704545	0.717391	0.436436	0.673913	0.760870
		SVM	0.695652	0.674419	0.695652	0.394676	0.630435	0.760870
		RF	0.684783	0.707071	0.684783	0.373920	0.760870	0.608696
	UMAP	KNN	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
		MLP	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		SVM	0.826087	0.836735	0.826087	0.657794	0.891304	0.760870
		RF	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
	Ivis	KNN	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
		RF	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
21	PCA	KNN	0.641304	0.697248	0.641304	0.304140	0.826087	0.456522
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.695652	0.735849	0.695652	0.410792	0.847826	0.543478
		RF	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
	TSVD	KNN	0.597826	0.633663	0.597826	0.199508	0.695652	0.500000
		MLP	0.706522	0.765217	0.706522	0.476942	0.956522	0.456522
		SVM	0.706522	0.696629	0.706522	0.413925	0.673913	0.739130
		RF	0.739130	0.769231	0.739130	0.495415	0.869565	0.608696
	KPCA	KNN	0.663043	0.710280	0.663043	0.344942	0.826087	0.500000
		MLP	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
		SVM	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		RF	0.684783	0.701031	0.684783	0.371768	0.739130	0.630435
	UMAP	KNN	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		MLP	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		SVM	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
	Ivis	KNN	0.750000	0.776699	0.750000	0.514940	0.869565	0.630435
		MLP	0.782609	0.772727	0.782609	0.567367	0.739130	0.826087
		SVM	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
		RF	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
22	PCA	KNN	0.630435	0.673077	0.630435	0.270226	0.760870	0.500000
		MLP	0.739130	0.733333	0.739130	0.478714	0.717391	0.760870

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		SVM	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
		RF	0.771739	0.783505	0.771739	0.546718	0.826087	0.717391
	TSVD	KNN	0.630435	0.666667	0.630435	0.267261	0.739130	0.521739
		MLP	0.641304	0.659794	0.641304	0.284293	0.695652	0.586957
		SVM	0.739130	0.744681	0.739130	0.478714	0.760870	0.717391
		RF	0.739130	0.769231	0.739130	0.495415	0.869565	0.608696
	KPCA	KNN	0.771739	0.796117	0.771739	0.559717	0.891304	0.652174
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.673913	0.693878	0.673913	0.350823	0.739130	0.608696
		RF	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
	UMAP	KNN	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		MLP	0.706522	0.737864	0.706522	0.425385	0.826087	0.586957
		SVM	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
		RF	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
	Ivis	KNN	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
		RF	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
23	PCA	KNN	0.630435	0.673077	0.630435	0.270226	0.760870	0.500000
		MLP	0.695652	0.745455	0.695652	0.425210	0.891304	0.500000
		SVM	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		RF	0.793478	0.815534	0.793478	0.604494	0.913043	0.673913
	TSVD	KNN	0.619565	0.666667	0.619565	0.249293	0.760870	0.478261
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.760870	0.765957	0.760870	0.522233	0.782609	0.739130
		RF	0.717391	0.750000	0.717391	0.450377	0.847826	0.586957
	KPCA	KNN	0.750000	0.780952	0.750000	0.521249	0.891304	0.608696
		MLP	0.739130	0.769231	0.739130	0.495415	0.869565	0.608696
		SVM	0.706522	0.727273	0.706522	0.417911	0.782609	0.630435
		RF	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
	UMAP	KNN	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		MLP	0.717391	0.723404	0.717391	0.435194	0.739130	0.695652
		SVM	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		RF	0.750000	0.752688	0.750000	0.500118	0.760870	0.739130
	Ivis	KNN	0.750000	0.776699	0.750000	0.514940	0.869565	0.630435
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.717391	0.754717	0.717391	0.456435	0.869565	0.565217
		RF	0.750000	0.752688	0.750000	0.500118	0.760870	0.739130
24	PCA	KNN	0.630435	0.679245	0.630435	0.273861	0.782609	0.478261
		MLP	0.728261	0.719101	0.728261	0.457496	0.695652	0.760870
		SVM	0.750000	0.780952	0.750000	0.521249	0.891304	0.608696
		RF	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
	TSVD	KNN	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696
		RF	0.782609	0.803922	0.782609	0.579066	0.891304	0.673913
	KPCA	KNN	0.750000	0.792793	0.750000	0.549021	0.956522	0.543478

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		MLP	0.706522	0.721649	0.706522	0.415505	0.760870	0.652174
		SVM	0.652174	0.680000	0.652174	0.309058	0.739130	0.565217
		RF	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
	UMAP	KNN	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
		MLP	0.706522	0.721649	0.706522	0.415505	0.760870	0.652174
		SVM	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		RF	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
	Ivis	KNN	0.695652	0.730769	0.695652	0.405340	0.826087	0.565217
		MLP	0.771739	0.796117	0.771739	0.559717	0.891304	0.652174
		SVM	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		RF	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
25	PCA	KNN	0.619565	0.666667	0.619565	0.249293	0.760870	0.478261
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
		RF	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
	TSVD	KNN	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.771739	0.774194	0.771739	0.543607	0.782609	0.760870
		RF	0.760870	0.788462	0.760870	0.540453	0.891304	0.630435
	KPCA	KNN	0.673913	0.693878	0.673913	0.350823	0.739130	0.608696
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.706522	0.737864	0.706522	0.425385	0.826087	0.586957
		RF	0.771739	0.783505	0.771739	0.546718	0.826087	0.717391
	UMAP	KNN	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		MLP	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
		SVM	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		RF	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
	Ivis	KNN	0.684783	0.712871	0.684783	0.376848	0.782609	0.586957
		MLP	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		SVM	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
		RF	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
26	PCA	KNN	0.641304	0.685714	0.641304	0.294619	0.782609	0.500000
		MLP	0.706522	0.742857	0.706522	0.430597	0.847826	0.565217
		SVM	0.782609	0.787234	0.782609	0.565752	0.804348	0.760870
		RF	0.782609	0.791667	0.782609	0.567367	0.826087	0.739130
	TSVD	KNN	0.706522	0.742857	0.706522	0.430597	0.847826	0.565217
		MLP	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		SVM	0.771739	0.778947	0.771739	0.544638	0.804348	0.739130
		RF	0.695652	0.730769	0.695652	0.405340	0.826087	0.565217
	KPCA	KNN	0.652174	0.680000	0.652174	0.309058	0.739130	0.565217
		MLP	0.717391	0.690476	0.717391	0.441511	0.630435	0.804348
		SVM	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		RF	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
	UMAP	KNN	0.771739	0.796117	0.771739	0.559717	0.891304	0.652174
		MLP	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		SVM	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
		RF	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
	Ivis	KNN	0.717391	0.750000	0.717391	0.450377	0.847826	0.586957
		MLP	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
		SVM	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
		RF	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
27	PCA	KNN	0.619565	0.672897	0.619565	0.252957	0.782609	0.456522
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		RF	0.782609	0.791667	0.782609	0.567367	0.826087	0.739130
	TSVD	KNN	0.608696	0.617021	0.608696	0.217597	0.630435	0.586957
		MLP	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
		SVM	0.782609	0.787234	0.782609	0.565752	0.804348	0.760870
		RF	0.750000	0.780952	0.750000	0.521249	0.891304	0.608696
	KPCA	KNN	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
		RF	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
	UMAP	KNN	0.706522	0.727273	0.706522	0.417911	0.782609	0.630435
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		RF	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
	Ivis	KNN	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		MLP	0.804348	0.816327	0.804348	0.613941	0.869565	0.739130
		SVM	0.695652	0.720000	0.695652	0.397360	0.782609	0.608696
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
28	PCA	KNN	0.695652	0.740741	0.695652	0.417365	0.869565	0.521739
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
		RF	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
	TSVD	KNN	0.706522	0.742857	0.706522	0.430597	0.847826	0.565217
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.760870	0.765957	0.760870	0.522233	0.782609	0.739130
		RF	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
	KPCA	KNN	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
		MLP	0.793478	0.804124	0.793478	0.590455	0.847826	0.739130
		SVM	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
		RF	0.760870	0.765957	0.760870	0.522233	0.782609	0.739130
	UMAP	KNN	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
		MLP	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		SVM	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
		RF	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
	Ivis	KNN	0.728261	0.761905	0.728261	0.475923	0.869565	0.586957
		MLP	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
		SVM	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		RF	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
29	PCA	KNN	0.706522	0.752294	0.706522	0.444513	0.891304	0.521739
		MLP	0.510870	0.042553	0.510870	0.104828	0.021739	1.000000
		SVM	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		RF	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
	TSVD	KNN	0.630435	0.673077	0.630435	0.270226	0.760870	0.500000
		MLP	0.728261	0.719101	0.728261	0.457496	0.695652	0.760870
		SVM	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		RF	0.739130	0.764706	0.739130	0.489979	0.847826	0.630435
	KPCA	KNN	0.739130	0.777778	0.739130	0.510113	0.913043	0.565217
		MLP	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		SVM	0.695652	0.720000	0.695652	0.397360	0.782609	0.608696
		RF	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
	UMAP	KNN	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		RF	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
	Ivis	KNN	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
		MLP	0.760870	0.784314	0.760870	0.534522	0.869565	0.652174
		SVM	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
		RF	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
30	PCA	KNN	0.717391	0.767857	0.717391	0.482805	0.934783	0.500000
		MLP	0.717391	0.754717	0.717391	0.456435	0.869565	0.565217
		SVM	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
		RF	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
	TSVD	KNN	0.695652	0.730769	0.695652	0.405340	0.826087	0.565217
		MLP	0.673913	0.571429	0.673913	0.396059	0.434783	0.913043
		SVM	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
		RF	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
	KPCA	KNN	0.663043	0.680412	0.663043	0.328031	0.717391	0.608696
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.673913	0.693878	0.673913	0.350823	0.739130	0.608696
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
	UMAP	KNN	0.706522	0.727273	0.706522	0.417911	0.782609	0.630435
		MLP	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		SVM	0.695652	0.740741	0.695652	0.417365	0.869565	0.521739
		RF	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
	Ivis	KNN	0.695652	0.720000	0.695652	0.397360	0.782609	0.608696
		MLP	0.750000	0.729412	0.750000	0.505892	0.673913	0.826087
		SVM	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		RF	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
31	PCA	KNN	0.695652	0.740741	0.695652	0.417365	0.869565	0.521739
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		RF	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
	TSVD	KNN	0.663043	0.704762	0.663043	0.339945	0.804348	0.521739
		MLP	0.728261	0.719101	0.728261	0.457496	0.695652	0.760870
		SVM	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
		RF	0.771739	0.783505	0.771739	0.546718	0.826087	0.717391
	KPCA	KNN	0.706522	0.727273	0.706522	0.417911	0.782609	0.630435
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		SVM	0.717391	0.717391	0.717391	0.434783	0.717391	0.717391
		RF	0.750000	0.776699	0.750000	0.514940	0.869565	0.630435
	UMAP	KNN	0.717391	0.750000	0.717391	0.450377	0.847826	0.586957
		MLP	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
		SVM	0.728261	0.766355	0.728261	0.482918	0.891304	0.565217
		RF	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
	Ivis	KNN	0.706522	0.737864	0.706522	0.425385	0.826087	0.586957
		MLP	0.695652	0.735849	0.695652	0.410792	0.847826	0.543478
		SVM	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		RF	0.771739	0.783505	0.771739	0.546718	0.826087	0.717391
32	PCA	KNN	0.695652	0.740741	0.695652	0.417365	0.869565	0.521739
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
	TSVD	KNN	0.652174	0.692308	0.652174	0.315264	0.782609	0.521739
		MLP	0.739130	0.733333	0.739130	0.478714	0.717391	0.760870
		SVM	0.706522	0.721649	0.706522	0.415505	0.760870	0.652174
		RF	0.793478	0.804124	0.793478	0.590455	0.847826	0.739130
	KPCA	KNN	0.750000	0.785047	0.750000	0.528910	0.913043	0.586957
		MLP	0.760870	0.788462	0.760870	0.540453	0.891304	0.630435
		SVM	0.684783	0.688172	0.684783	0.369653	0.695652	0.673913
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
	UMAP	KNN	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		MLP	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		SVM	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		RF	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
	Ivis	KNN	0.728261	0.766355	0.728261	0.482918	0.891304	0.565217
		MLP	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		SVM	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
		RF	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
33	PCA	KNN	0.695652	0.740741	0.695652	0.417365	0.869565	0.521739
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
	TSVD	KNN	0.673913	0.711538	0.673913	0.360302	0.804348	0.543478
		MLP	0.750000	0.796460	0.750000	0.561979	0.978261	0.521739
		SVM	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		RF	0.793478	0.808081	0.793478	0.593873	0.869565	0.717391
	KPCA	KNN	0.739130	0.781818	0.739130	0.519701	0.934783	0.543478
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.695652	0.708333	0.695652	0.392792	0.739130	0.652174
		RF	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
	UMAP	KNN	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		MLP	0.663043	0.680412	0.663043	0.328031	0.717391	0.608696
		SVM	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		RF	0.717391	0.723404	0.717391	0.435194	0.739130	0.695652
	Ivis	KNN	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
		RF	0.793478	0.800000	0.793478	0.588209	0.826087	0.760870
34	PCA	KNN	0.706522	0.747664	0.706522	0.436926	0.869565	0.543478
		MLP	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
		SVM	0.739130	0.773585	0.739130	0.502079	0.891304	0.586957
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
	TSVD	KNN	0.695652	0.735849	0.695652	0.410792	0.847826	0.543478
		MLP	0.739130	0.739130	0.739130	0.478261	0.739130	0.739130
		SVM	0.739130	0.769231	0.739130	0.495415	0.869565	0.608696
		RF	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
	KPCA	KNN	0.717391	0.754717	0.717391	0.456435	0.869565	0.565217
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.750000	0.780952	0.750000	0.521249	0.891304	0.608696
		RF	0.793478	0.811881	0.793478	0.598524	0.891304	0.695652
	UMAP	KNN	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		MLP	0.706522	0.756757	0.706522	0.453539	0.913043	0.500000
		SVM	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		RF	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
	Ivis	KNN	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		RF	0.804348	0.820000	0.804348	0.618115	0.891304	0.717391
35	PCA	KNN	0.586957	0.604167	0.586957	0.174574	0.630435	0.543478
		MLP	0.728261	0.705882	0.728261	0.461901	0.652174	0.804348
		SVM	0.793478	0.811881	0.793478	0.598524	0.891304	0.695652
		RF	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
	TSVD	KNN	0.630435	0.679245	0.630435	0.273861	0.782609	0.478261
		MLP	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		SVM	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		RF	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
	KPCA	KNN	0.695652	0.702128	0.695652	0.391675	0.717391	0.673913
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
		RF	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
	UMAP	KNN	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		MLP	0.652174	0.652174	0.652174	0.304348	0.652174	0.652174
		SVM	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
		RF	0.706522	0.721649	0.706522	0.415505	0.760870	0.652174
	Ivis	KNN	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		MLP	0.728261	0.761905	0.728261	0.475923	0.869565	0.586957
		SVM	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
		RF	0.739130	0.764706	0.739130	0.489979	0.847826	0.630435
36	PCA	KNN	0.663043	0.699029	0.663043	0.335830	0.782609	0.543478
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.760870	0.784314	0.760870	0.534522	0.869565	0.652174
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
	TSVD	KNN	0.630435	0.690909	0.630435	0.283473	0.826087	0.434783
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.717391	0.754717	0.717391	0.456435	0.869565	0.565217
		RF	0.760870	0.784314	0.760870	0.534522	0.869565	0.652174
	KPCA	KNN	0.760870	0.800000	0.760870	0.566947	0.956522	0.565217
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		RF	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
	UMAP	KNN	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		MLP	0.717391	0.763636	0.717391	0.472456	0.913043	0.521739
		SVM	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		RF	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
	Ivis	KNN	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
37	PCA	KNN	0.673913	0.705882	0.673913	0.356348	0.782609	0.565217
		MLP	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		SVM	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		RF	0.771739	0.783505	0.771739	0.546718	0.826087	0.717391
	TSVD	KNN	0.673913	0.711538	0.673913	0.360302	0.804348	0.543478
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
		RF	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
	KPCA	KNN	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.760870	0.784314	0.760870	0.534522	0.869565	0.652174
		RF	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
	UMAP	KNN	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		MLP	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		SVM	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		RF	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
	Ivis	KNN	0.739130	0.769231	0.739130	0.495415	0.869565	0.608696
		MLP	0.695652	0.725490	0.695652	0.400892	0.804348	0.586957
		SVM	0.760870	0.760870	0.760870	0.521739	0.760870	0.760870
		RF	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
38	PCA	KNN	0.663043	0.710280	0.663043	0.344942	0.826087	0.500000
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		RF	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
	TSVD	KNN	0.652174	0.680000	0.652174	0.309058	0.739130	0.565217
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
		RF	0.760870	0.784314	0.760870	0.534522	0.869565	0.652174
	KPCA	KNN	0.684783	0.694737	0.684783	0.370354	0.717391	0.652174
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		RF	0.673913	0.693878	0.673913	0.350823	0.739130	0.608696
	UMAP	KNN	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		MLP	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
		SVM	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		RF	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
	Ivis	KNN	0.750000	0.780952	0.750000	0.521249	0.891304	0.608696
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
		RF	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
39	PCA	KNN	0.663043	0.693069	0.663043	0.332513	0.760870	0.565217
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
	TSVD	KNN	0.695652	0.750000	0.695652	0.434524	0.913043	0.478261
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.771739	0.783505	0.771739	0.546718	0.826087	0.717391
		RF	0.782609	0.803922	0.782609	0.579066	0.891304	0.673913
	KPCA	KNN	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		RF	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
	UMAP	KNN	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
		MLP	0.706522	0.709677	0.706522	0.413141	0.717391	0.695652
		SVM	0.728261	0.742268	0.728261	0.459243	0.782609	0.673913
		RF	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
	Ivis	KNN	0.706522	0.737864	0.706522	0.425385	0.826087	0.586957
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.750000	0.780952	0.750000	0.521249	0.891304	0.608696
		RF	0.782609	0.803922	0.782609	0.579066	0.891304	0.673913
40	PCA	KNN	0.652174	0.703704	0.652174	0.324617	0.826087	0.478261
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
	TSVD	KNN	0.673913	0.727273	0.673913	0.377964	0.869565	0.478261
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
		RF	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
	KPCA	KNN	0.750000	0.785047	0.750000	0.528910	0.913043	0.586957
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		RF	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
	UMAP	KNN	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		RF	0.728261	0.747475	0.728261	0.461901	0.804348	0.652174
	Ivis	KNN	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		SVM	0.717391	0.723404	0.717391	0.435194	0.739130	0.695652
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
41	PCA	KNN	0.630435	0.666667	0.630435	0.267261	0.739130	0.521739
		MLP	0.793478	0.808081	0.793478	0.593873	0.869565	0.717391
		SVM	0.750000	0.776699	0.750000	0.514940	0.869565	0.630435
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
	TSVD	KNN	0.695652	0.740741	0.695652	0.417365	0.869565	0.521739
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.739130	0.744681	0.739130	0.478714	0.760870	0.717391
		RF	0.804348	0.816327	0.804348	0.613941	0.869565	0.739130
	KPCA	KNN	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		RF	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
	UMAP	KNN	0.760870	0.784314	0.760870	0.534522	0.869565	0.652174
		MLP	0.684783	0.718447	0.684783	0.380608	0.804348	0.565217
		SVM	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		RF	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
	Ivis	KNN	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		MLP	0.695652	0.740741	0.695652	0.417365	0.869565	0.521739
		SVM	0.771739	0.800000	0.771739	0.566575	0.913043	0.630435
		RF	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
42	PCA	KNN	0.641304	0.691589	0.641304	0.298949	0.804348	0.478261
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
	TSVD	KNN	0.641304	0.666667	0.641304	0.285939	0.717391	0.565217
		MLP	0.760870	0.755556	0.760870	0.522233	0.739130	0.782609
		SVM	0.739130	0.744681	0.739130	0.478714	0.760870	0.717391
		RF	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
	KPCA	KNN	0.684783	0.707071	0.684783	0.373920	0.760870	0.608696
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.793478	0.808081	0.793478	0.593873	0.869565	0.717391
		RF	0.760870	0.780000	0.760870	0.529813	0.847826	0.673913
	UMAP	KNN	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.695652	0.695652	0.695652	0.391304	0.695652	0.695652
		RF	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
	Ivis	KNN	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		MLP	0.739130	0.781818	0.739130	0.519701	0.934783	0.543478
		SVM	0.695652	0.708333	0.695652	0.392792	0.739130	0.652174
		RF	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
43	PCA	KNN	0.608696	0.647059	0.608696	0.222718	0.717391	0.500000
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
		RF	0.782609	0.791667	0.782609	0.567367	0.826087	0.739130
	TSVD	KNN	0.684783	0.712871	0.684783	0.376848	0.782609	0.586957

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.815217	0.828283	0.815217	0.637864	0.891304	0.739130
		RF	0.782609	0.791667	0.782609	0.567367	0.826087	0.739130
	KPCA	KNN	0.684783	0.707071	0.684783	0.373920	0.760870	0.608696
		MLP	0.804348	0.820000	0.804348	0.618115	0.891304	0.717391
		SVM	0.804348	0.816327	0.804348	0.613941	0.869565	0.739130
		RF	0.771739	0.783505	0.771739	0.546718	0.826087	0.717391
	UMAP	KNN	0.695652	0.725490	0.695652	0.400892	0.804348	0.586957
		MLP	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		SVM	0.684783	0.688172	0.684783	0.369653	0.695652	0.673913
		RF	0.750000	0.752688	0.750000	0.500118	0.760870	0.739130
	Ivis	KNN	0.739130	0.764706	0.739130	0.489979	0.847826	0.630435
		MLP	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
		SVM	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
		RF	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
44	PCA	KNN	0.641304	0.702703	0.641304	0.310316	0.847826	0.434783
		MLP	0.782609	0.803922	0.782609	0.579066	0.891304	0.673913
		SVM	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
	TSVD	KNN	0.608696	0.672727	0.608696	0.236228	0.804348	0.413043
		MLP	0.793478	0.795699	0.793478	0.587095	0.804348	0.782609
		SVM	0.793478	0.811881	0.793478	0.598524	0.891304	0.695652
		RF	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
	KPCA	KNN	0.684783	0.712871	0.684783	0.376848	0.782609	0.586957
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
		RF	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
	UMAP	KNN	0.717391	0.745098	0.717391	0.445435	0.826087	0.608696
		MLP	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		SVM	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		RF	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
	Ivis	KNN	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		MLP	0.728261	0.761905	0.728261	0.475923	0.869565	0.586957
		SVM	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
45	PCA	KNN	0.641304	0.702703	0.641304	0.310316	0.847826	0.434783
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
	TSVD	KNN	0.608696	0.672727	0.608696	0.236228	0.804348	0.413043
		MLP	0.804348	0.816327	0.804348	0.613941	0.869565	0.739130
		SVM	0.804348	0.820000	0.804348	0.618115	0.891304	0.717391
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
	KPCA	KNN	0.706522	0.732673	0.706522	0.421184	0.804348	0.608696
		MLP	0.804348	0.816327	0.804348	0.613941	0.869565	0.739130
		SVM	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		RF	0.771739	0.778947	0.771739	0.544638	0.804348	0.739130

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
	UMAP	KNN	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
		MLP	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		SVM	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		RF	0.760870	0.775510	0.760870	0.526235	0.826087	0.695652
	Ivis	KNN	0.739130	0.760000	0.739130	0.485662	0.826087	0.652174
		MLP	0.750000	0.762887	0.750000	0.502980	0.804348	0.695652
		SVM	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
46	PCA	KNN	0.641304	0.679612	0.641304	0.291053	0.760870	0.521739
		MLP	0.750000	0.792793	0.750000	0.549021	0.956522	0.543478
		SVM	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
		RF	0.760870	0.770833	0.760870	0.523723	0.804348	0.717391
	TSVD	KNN	0.663043	0.710280	0.663043	0.344942	0.826087	0.500000
		MLP	0.771739	0.787879	0.771739	0.549882	0.847826	0.695652
		SVM	0.804348	0.820000	0.804348	0.618115	0.891304	0.717391
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
	KPCA	KNN	0.717391	0.754717	0.717391	0.456435	0.869565	0.565217
		MLP	0.739130	0.764706	0.739130	0.489979	0.847826	0.630435
		SVM	0.804348	0.816327	0.804348	0.613941	0.869565	0.739130
		RF	0.760870	0.770833	0.760870	0.523723	0.804348	0.717391
	UMAP	KNN	0.717391	0.740000	0.717391	0.441511	0.804348	0.630435
		MLP	0.695652	0.708333	0.695652	0.392792	0.739130	0.652174
		SVM	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
	Ivis	KNN	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		MLP	0.750000	0.780952	0.750000	0.521249	0.891304	0.608696
		SVM	0.760870	0.765957	0.760870	0.522233	0.782609	0.739130
		RF	0.695652	0.714286	0.695652	0.394676	0.760870	0.630435
47	PCA	KNN	0.652174	0.686275	0.652174	0.311805	0.760870	0.543478
		MLP	0.815217	0.828283	0.815217	0.637864	0.891304	0.739130
		SVM	0.804348	0.816327	0.804348	0.613941	0.869565	0.739130
		RF	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
	TSVD	KNN	0.673913	0.716981	0.673913	0.365148	0.826087	0.521739
		MLP	0.500000	0.666667	0.500000	0.000000	1.000000	0.000000
		SVM	0.804348	0.820000	0.804348	0.618115	0.891304	0.717391
		RF	0.750000	0.757895	0.750000	0.501067	0.782609	0.717391
	KPCA	KNN	0.706522	0.737864	0.706522	0.425385	0.826087	0.586957
		MLP	0.826087	0.840000	0.826087	0.662266	0.913043	0.739130
		SVM	0.793478	0.808081	0.793478	0.593873	0.869565	0.717391
		RF	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
	UMAP	KNN	0.728261	0.757282	0.728261	0.470162	0.847826	0.608696
		MLP	0.717391	0.759259	0.717391	0.463739	0.891304	0.543478
		SVM	0.739130	0.750000	0.739130	0.480079	0.782609	0.695652
		RF	0.706522	0.727273	0.706522	0.417911	0.782609	0.630435
	Ivis	KNN	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.793478	0.786517	0.793478	0.588209	0.760870	0.826087

Continued on next page

Table 5 – Independent test results for the best models found.

		Measure	Accuracy	F1	AUC	MCC	Sensitivity	Specificity
Dimensionality	Projector	Classifier						
		RF	0.673913	0.687500	0.673913	0.349149	0.717391	0.630435
48	PCA	KNN	0.673913	0.711538	0.673913	0.360302	0.804348	0.543478
		MLP	0.836957	0.842105	0.836957	0.675351	0.869565	0.804348
		SVM	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
		RF	0.782609	0.795918	0.782609	0.570088	0.847826	0.717391
	TSVD	KNN	0.673913	0.716981	0.673913	0.365148	0.826087	0.521739
		MLP	0.836957	0.838710	0.836957	0.674072	0.847826	0.826087
		SVM	0.826087	0.840000	0.826087	0.662266	0.913043	0.739130
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
	KPCA	KNN	0.673913	0.700000	0.673913	0.353209	0.760870	0.586957
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.815217	0.828283	0.815217	0.637864	0.891304	0.739130
		RF	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
	UMAP	KNN	0.717391	0.729167	0.717391	0.436436	0.760870	0.673913
		MLP	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		SVM	0.750000	0.776699	0.750000	0.514940	0.869565	0.630435
		RF	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
	Ivis	KNN	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
		MLP	0.760870	0.792453	0.760870	0.547723	0.913043	0.608696
		SVM	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
		RF	0.750000	0.767677	0.750000	0.505892	0.826087	0.673913
49	PCA	KNN	0.673913	0.711538	0.673913	0.360302	0.804348	0.543478
		MLP	0.739130	0.793103	0.739130	0.560612	1.000000	0.478261
		SVM	0.793478	0.808081	0.793478	0.593873	0.869565	0.717391
		RF	0.739130	0.755102	0.739130	0.482382	0.804348	0.673913
	TSVD	KNN	0.706522	0.747664	0.706522	0.436926	0.869565	0.543478
		MLP	0.500000	0.000000	0.500000	0.000000	0.000000	1.000000
		SVM	0.815217	0.828283	0.815217	0.637864	0.891304	0.739130
		RF	0.771739	0.792079	0.771739	0.554189	0.869565	0.673913
	KPCA	KNN	0.673913	0.700000	0.673913	0.353209	0.760870	0.586957
		MLP	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652
		SVM	0.793478	0.808081	0.793478	0.593873	0.869565	0.717391
		RF	0.750000	0.772277	0.750000	0.509854	0.847826	0.652174
	UMAP	KNN	0.695652	0.720000	0.695652	0.397360	0.782609	0.608696
		MLP	0.728261	0.736842	0.728261	0.457496	0.760870	0.695652
		SVM	0.717391	0.734694	0.717391	0.438529	0.782609	0.652174
		RF	0.706522	0.709677	0.706522	0.413141	0.717391	0.695652
	Ivis	KNN	0.728261	0.752475	0.728261	0.465519	0.826087	0.630435
		MLP	0.793478	0.811881	0.793478	0.598524	0.891304	0.695652
		SVM	0.739130	0.764706	0.739130	0.489979	0.847826	0.630435
		RF	0.782609	0.800000	0.782609	0.573964	0.869565	0.695652