|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Data preprocess** | **Feature engineering** | **Hyper parameter tuning** | **Model training** | **Evaluation metrics** | **Model interpretability** | **Link** |
|  | The BERT is used to extract features, namely word and sentence embedding vectors, from text data. these vectors are used as high-quality feature inputs to downstream models. NLP models such as LSTMs or CNNs require inputs in the form of numerical vectors, and this typically means translating features like the vocabulary and parts of speech into numerical representations. |  | The accuracy of the BERT model would have likely further increased if we retrained the BERT model on research literature instead of using a generic model trained on Wikipedia. |  |  | <https://link.springer.com/article/10.1007/s11192-022-04314-9> |
|  | The application of the content based filter allows us to pass from the field of images to that of numbers, thus transforming a data set of images into a numerical one. This latter can be easily deal with evolutionary algorithm. |  |  |  |  | <https://link.springer.com/article/10.1007/s00521-021-06806-w> |
| models using only dichotomous features perform only slightly worse than models based on a complex combination of numerical input values |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S2352914823000308> |
|  |  |  | while logistic regression summarizes the influence of each feature with a single coefficient and limited to clustering patients, or regression trees, GAMs accommodate non-linear and non montonic effects |  |  | <https://www.sciencedirect.com/science/article/pii/S1532046422001022> |
|  |  |  |  |  | RIPPER is a rule-based ML algorithm, in which rules are learned from the data directly. pyFUME can create rules based on fuzzy logic. the potential of pyFUME and RIPPER in providing inherent interpretability are reflected. | <https://ieeexplore.ieee.org/abstract/document/9863020> |
|  |  |  |  |  | GNAMs belong to the GAMs family and learn a linear combination of multi-layer perceptron (MLP) with an input, an output, and several hidden layers. | <https://link.springer.com/article/10.1186/s12874-022-01827-y> |
|  | Before running the machine learning models, a descriptive correlation statistics analysis is carried out comprising the p-values for the predictor variables and the target variable in the original datasets. |  |  |  |  | <https://link.springer.com/article/10.1186/s12889-023-16236-z> |
|  | The pipeline mines over 60 thousand raw (e.g., diagnosis, medication and procedure codes) and 160 million transitive sequential (e.g., medication to diagnosis) representations. Through the iterative feature selection by means of MSMR flter method and embedded methods, these features were shrunk to over 2300 representations, about 900 of which were raw features and about 1400 were transitive sequential features. |  |  |  |  | <https://www.nature.com/articles/s41598-021-84781-x> |
|  |  |  | SIMLIR is an interpretable probabilistic graphical model (PGM) that utilies ML inside the compartmental SIR model (the changes in the policies implemented at the government level is used to estimate the time-varying parameters of an SIR model) to explicitly incorporate government policy decisions |  |  | <https://www.mdpi.com/2571-9394/4/1/5> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9179729> |
| synthetic data is generated using the VAE model. |  |  |  |  |  | <https://www.nature.com/articles/s41598-023-31542-7> |
|  |  |  |  |  |  | <https://www.mdpi.com/2075-1680/11/8/374> |
|  |  |  | xgb is an ensemble model and creates multiple weak classifiers which work strongly when together. |  |  | <https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000020> |
|  |  |  |  |  | EBM is constructed with multiple hierarchically organized simple classifiers consisting of sequences of binary decisions. Unlike the black-box models, EBM produce lossless explanations for outcome predictions due to its great interpretability potential of tree-based decision system. | <https://www.medrxiv.org/content/10.1101/2020.06.30.20143651v1> |
|  |  |  |  |  | Temporal  Fusion Transformer TFT adds a kind of interpretable multi-head self-attention mechanism to lstm to better learn the different temporal patterns. (2) attention weights provides insights on feature importances. (3) the sensitivity of a model input is the ratio of the change in an output variable to the change in an input feature. (4) The conceptualization of modeling individual features with the post hoc sensitivity analysis can apply to other time series models as the Morris method is efficient and generic. | <https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/ru/ppzbmed-2210.03258v1> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9298069?casa_token=m6oiKIyCDvoAAAAA:U0bIZU7qphPaabOjBmXIQH0YTbCuJ0e6hFGRgYQyQW9YMJBfmb9uWAIiBz75SIzAgZKjJ0Mel1k> |
|  | The factor analysis technique contributes to identify latent composite variables. Factor Analysis accounts for the fact that multiple socio-economic factors may be strongly correlated, for example between GDPs per capita and other development metrics e.g., access to electricity to underlying latent factors which are represented by other variables. |  |  |  | One of the strengths of EBM is its ability to consider many factors simultaneously and inform their relative impacts on the dependent variables. This is comparable to the factor analysis, but instead of aggregating correlating factors in the same category to construct latent variables, EBM highlights the most representative factor. | <https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/ru/ppzbmed-10.21203.rs.3.rs-2877405.v1> |
|  |  |  |  |  |  | <https://www.medrxiv.org/content/10.1101/2021.01.04.21249218v1> |
| a sequential feature selection method works as follows: It removes features one by one from the full set of features and evaluates the error function. When the error reaches the optimumlevel, the combination of left-over features is regarded as the optimal feature combination. . (2) importance of each selected factor determined by xgb feature importance (3) SHAP analysis was used to calculate the contribution of selected factors for each city. |  |  |  |  |  | <https://www.jmir.org/2020/11/e23853/> |
|  |  |  | xgb is frequently reported as a robust method in the context of COVID-19 research. |  |  | <https://www.sciencedirect.com/science/article/pii/S0010482523000847> |
|  |  |  |  |  |  | <https://id.elsevier.com/as/authorization.oauth2?platSite=SD%2Fscience&scope=openid%20email%20profile%20els_auth_info%20els_idp_info%20els_idp_analytics_attrs%20urn%3Acom%3Aelsevier%3Aidp%3Apolicy%3Aproduct%3Ainst_assoc&response_type=code&redirect_uri=https%3A%2F%2Fwww.sciencedirect.com%2Fuser%2Fidentity%2Flanding&authType=SINGLE_SIGN_IN&prompt=none&client_id=SDFE-v4&state=retryCounter%3D0%26csrfToken%3De3cb4bc2-d519-432c-ae4f-3e3c5def987c%26idpPolicy%3Durn%253Acom%253Aelsevier%253Aidp%253Apolicy%253Aproduct%253Ainst_assoc%26returnUrl%3D%252Fscience%252Farticle%252Fpii%252FS0010482521001293%26prompt%3Dnone%26cid%3Darp-9dc1627e-ac2a-4b03-bf2b-a730cad9c2ce> |
|  |  |  |  |  | causal Shapley values account for subjective or pre knowledge based graphical causal structure of the data into account in when estimating the Shapley values. | <https://www.nature.com/articles/s41598-022-18725-4> |
|  |  |  |  | The ann feature selcection's objective is to maximize AUC score and minimize the number of features selected for model development. |  | <https://www.frontiersin.org/articles/10.3389/fpubh.2021.626697/full> |
|  |  |  |  |  | Through Formal Concept Analysis (FCA) approach, a set of association rules are generated with different confidence intervalls, which comprise clear implications between the attributes in the formal context. | <https://www.sciencedirect.com/science/article/pii/S0933365722001452?casa_token=xZ2ZfpJF4BIAAAAA:3t4UDDZyoGcfcwSUdcVX0Ibs1LsQsR6QtG0AMf7gtuo77G6ghaDRaWQoaSIPadVXnzLyOiaQ> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S1076633222002008> |
|  |  |  |  |  | a hybrid architecture that combines autoregressive AR and LSTM models additively and trains the entire architecture jointly, allowing the relative weights of the interpretable AR part and uninterpretable predictive LSTM part to be fully determined by the data. | <https://www.nature.com/articles/s41598-023-33685-z> |
|  |  |  |  |  |  | <https://link.springer.com/article/10.1007/s12559-020-09812-7> |
|  |  |  |  |  |  | <https://www.frontiersin.org/articles/10.3389/frai.2021.684609/full> |
|  |  |  |  |  | LIME perturbs the features in an example and fits a linear model to approximate the neural network at the local region in the feature space surrounding the example. It then uses the linear model to determine which features were most contributory to the model’s prediction for that example. | <https://towardsdatascience.com/investigation-of-explainable-predictions-of-covid-19-infection-from-chest-x-rays-with-machine-cb370f46af1d> |
|  |  |  |  |  |  | <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0240346> |
|  |  |  | Among non-linear classifiers, Gaussian process classifier is arguably one of the most powerful techniques with sound statistical properties. |  |  | <https://www.nature.com/articles/s41598-021-98071-z> |
|  |  |  |  |  | iCNN does not make assumptions about the relevance of local information. Instead, a model-free influence score (I-score) is used to directly extract the influential information from images to form important variable modules. This technique replaces all pretrained filters in well-documented cnns by trial-and-error with explainable, influential, and predictive variable sets (modules) determined by the I-score. the suggested algorithm identifies only the variables or pixels with high I-score values that are extremely predictive and important. | <https://www.mdpi.com/1999-4893/14/11/337> |
|  |  |  |  |  |  | <https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/fr/covidwho-1871934> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/10090012> |
|  | H-statistics indicates overall interaction strength of each feature with the other features. The interaction is defined as the share of variance that is explained by the interaction. , but it does not tell us how the interactions look like. That is what pdps are for. |  |  |  |  | <https://www.researchsquare.com/article/rs-52330/v1> |
| a convolutional neural network (CNN) layer before the transformer as a way to reduce the required size of the transformer heads |  |  |  |  | (3) The forward attention layer not only helps classify data but also allow us to visualize and potentially interpret the models that we train. | <https://www.researchsquare.com/article/rs-1234007/v1> |
|  |  |  |  |  | JRIP (rule classifier) showed the best performance in 10-fold cross validation, and the best average performance in a further validation test using a diferent patient dataset. Moreover, JRIP showed comparable performances with other approaches that do not provide a clear and/or understandable explanation. | <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-022-02076-1> |
|  |  |  |  |  |  | <https://link.springer.com/article/10.1007/s10994-020-05921-4> |
| (1) unlabeled data is exploited in augmenting the dataset to enhance the performance of the classifier , thus providing ample training data and overcoming the issue of label sparsity and label inconsistency. |  |  | each expert is modeled by a Deep Neural Network, and then the outputs of these expert models are combined to generate a final label for each sample |  |  | <https://www.sciencedirect.com/science/article/pii/S0169260723004091> |
|  | feature selction procedure consists of a voting system utilizing the feature coefficient metric for logistic regression and feature importance metric for gradient boosting and AdaBoost model |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0010482522001536?casa_token=2qkRXlZ5_t4AAAAA:U3evkD-v3275wrGgfpa-s79wt-5resKwppyh2tB6hFbcBq9JnVH5ppNomAH8axPSBRpcj5PI> |
|  |  |  |  |  |  | <https://www.future-science.com/doi/full/10.2144/fsoa-2020-0207> |
|  | (1) UMAP is a flexible non-linear dimension reduction algorithm to learn the manifold structure of the data and find a low dimensional embedding that preserves the essential topological structure of that manifold. (2) instead of applying clustering directly to feature data in an unsupervised manner, clustering is run on feature SHAP values (after xgboost classification). SHAP also serves as a pre-processing step that rescales feature data prior to clustering using Uniform Manifold Approximation and Projection (UMAP). |  |  |  | Finally, rather than characterise clusters using descriptive statistics, the paper uses the SkopeRules package to construct discriminative decision-rules that identify and differentiate the clusters, forming the subgroup descriptions. | <https://link.springer.com/chapter/10.1007/978-3-030-93733-1_29?xgtab=&%3B=&ref=aidan-coopers-blog> |
| ARIMA is found to be a suitable model for prediction COVID-19 cases in india, though, with little data, no good results were obtained, but having 4-month data the results obtained were promising. |  |  |  |  |  | <https://link.springer.com/chapter/10.1007/978-3-031-08637-3_2> |
|  |  |  | (0) bagging tries to reduce variance not bias but boosting tries to reduce bias not variance high variances are present(3) boosting: Boosting is a sequential ensemble method that iteratively adjusts the weight of observation as per the last classification e.g. adaboost |  |  | <https://www.igi-global.com/chapter/an-investigation-of-the-coronavirus-disease-covid-19-mortality-risk-using-machine-learning/301406> |
|  |  |  | (1) the nodes in gnn represent districts (2) the network data is augmented with spatial networks encoding neighboring districts and distances in kilometers (3)network data comprising colocations, social conectedness and people staying put, are incorporated into distric forecast. (5) The graph is the input, and each graphs component i.e. nodes edges get updated by a machine learning model e.g. based on the features networks of its neighbors and the edges connecting them. |  | (4) the genral framework fuses the interpretability of distributional regression with a GNN architecture | <https://arxiv.org/abs/2101.00661> |
|  | principal component analysis was used in the pretraining step to condense highly comorbid anxiety and depression symptoms into a single data-driven measure of emotional distress. |  |  |  |  | <https://mental.jmir.org/2021/11/e32876/> |
| For imputation, the multivariate k-nearest neighbors algorithm was used |  |  |  |  |  | <https://www.degruyter.com/document/doi/10.1515/cclm-2020-1294/html> |
|  |  |  | (1) the framework consists of a computational graph cg as an activity choice model based on the survey data (2) at the second step, based on the Google mobility dataset, the mobility demand change is predicted by applying the LSTM method (3) the utility function of choosing an activity is defined as the sum of a systematic portion for observed individual preferences and a constant portion that represents the impacts of some unobserved characteristics related to the activity choice. |  |  | <https://www.sciencedirect.com/science/article/pii/S0968090X21003636?casa_token=ANHkji2x9jsAAAAA:qMcn4lz3gZnioh9LjazIPBocEY8ULYXweC5AOD6JwS7WNhopL2GjCf0mMZlgp8rjwZsg355V> |
|  |  |  | A GBDT aggregates a large number of weak prediction models, in this case dts, into a robust prediction algorithm, where the presence of many trees mitigates the errors due to a single-tree prediction. |  | (1) The odds ratio is commonly used to report the strength of association between exposure and an event.(2) | <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009121> |
|  | (1)The DVAE model has inherent advantages being interpretable i.e. the input data can be encoded to some latent space features that reveal the hidden relationship among the input variables.(2) |  |  |  | Miller (2019) defined interpretability as: “Interpretability is the degree to which a human can understand the cause of a decision”. | <https://hammer.purdue.edu/ndownloader/files/31719740> |
|  |  |  | (1) CNN is capable of learning global and higher-level features through max-pooling from successive convolutional windows. However, CNN is not capable of sequential learning, while RNN deals with time-series and sequential data. |  |  | <https://www.researchgate.net/profile/Mohammad-Hossain-185/publication/361408466_Explainable_Text_Classification_Model_for_COVID-19_Fake_News_Detection/links/62aef70d40d84c1401b7e892/Explainable-Text-Classification-Model-for-COVID-19-Fake-News-Detection.pdf> |
| By trying to compare an unlabeled data point to the training dataset, the K-nearest-neighbor (KNN) classifier improves considerably. It finds the K most related data-points, which are termed as KNNs. A metric that measures distance such as Euclidean or Manhattan distance is widely utilized to determine proximity. This technique then assigns the given data point to the KNN's most familiar class. |  |  |  |  |  | <https://link.springer.com/article/10.1007/s12539-021-00499-4> |
|  |  |  |  |  |  | <https://link.springer.com/chapter/10.1007/978-3-031-08506-2_12> |
|  |  |  | (1) In this study, a logistic regression model was used as the meta-learner, and different combinations of SVM, XGB and ANNs were used as base-learners, resulting in four different instances of stacking ensemble models.(2) There are several ensemble learning techniques in literature however, the most common ones are: (a) Bagging, (b) Boosting, Stacking and Mixture of Experts. (3) the predictions of multiple models, referred to as base-models, are fed into a second-level model, referred to as a meta-learner. The meta-learner is then trained to optimally combine the predictions of base-learners, to form a final set of predictions. (4) e only discuss the feature importance of the linear models |  |  | <https://www.sciencedirect.com/science/article/pii/S0010482522007442> |
|  | (1) two different feature selection algorithms, including Bagging Random Forests (BRF) and Multivariate Adaptive Regression Splines (MARS) |  | (2) For multiclass classification, multinomial logistic regression (MLR) | Importance values (IVs) were calculated using generalized cross-validation criterion with normalization. |  | <https://www.nature.com/articles/s41598-022-18994-z> |
|  |  |  | (1) Two glass-box models, logistic regression and explainable boosting machine, and two black-box models, random forest and support vector machine, were used |  | (2) with local explanations the paper show that although the overall results are really good, the models still make case wise classification mistakes and get confused with the features | <https://link.springer.com/article/10.1007/s40313-021-00858-y> |
|  |  |  |  |  |  | <https://books.google.de/books?hl=en&lr=&id=I3lxEAAAQBAJ&oi=fnd&pg=PA5&dq=covid+intext:%22interpretable+machine+learning%22&ots=IaZ6OoM7wR&sig=7loT2hcNvtpO4RclQ_WvilOVglc#v=onepage&q=covid%20intext%3A%22interpretable%20machine%20learning%22&f=false> |
|  | The architecture of the proposed dCNNs model comprises an early feature extractor acting as candidate generator in an abstract feature space, followed by a discriminator sub-network used to compute probabilities on whether the abnormality is present or not (in an image sub-region of interest) |  |  |  |  | <https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-022-07617-7> |
|  | LASSO’s objective function is the same as in OLS regression, with the exception the LASSO solution also minimizes the sum of the absolute value of the model coefficients. The minimization of the absolute value of the predictor variable coefficient estimates helps penalize the model as it becomes more complex. When the shrinkage penalty λ gets sufficiently large, it will shrink coefficients down to zero, therefore removing variables from the model. Hence, it only keeps the variables that are really driving prediction and shrinks the rest down to zero. |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0739885923000161> |
|  | 1. Stepwise multiple regression analysis is a variable selection method 2. (3) We see that there is a high degree of correlation between multiple independent variables, such as F18 (Share of population with cancer) and F27 (Share who is 70 or over). The high correlation between these two variables is easy to understand because the risk of cancer is closely related to age and older age groups will show a higher risk of cancer. After the correlation analysis, we eliminate a total of 13 variables, including Share of population with cancer. |  | (2) the paper selects linear regression (LR), support vector regression (SVR), and extreme learning machine (ELM) as the base learners. At the same time, it uses linear regression as the combiner of the second layer of the model. |  |  | <https://www.sciencedirect.com/science/article/pii/S1568494621008681?casa_token=awLgwwPr1cwAAAAA:tNRPcz3-WUgepJOIajcAETANlYqkoonAVTsbBXaaQd5Chs9eY8EzbDPsPKEiq-v4UjC2POMj> |
|  |  |  |  |  |  | <https://www.jmir.org/2021/4/e26211/> |
|  |  |  | (1) The belief rule based expert system can integrate the advantages of expert knowledge based plus data-driven approaches in a single framework. The initial belief rule base is established by expert knowledge which is further optimized by COVID-19 patients’ data using a modified differential evolution algorithm. |  |  | <https://www.mdpi.com/2076-3417/11/13/5810> |
|  |  |  |  |  |  | <https://www.proquest.com/openview/2182a86de3e6939add0a47ca718cf388/1?pq-origsite=gscholar&cbl=4986839&casa_token=D4EoMEsyuYoAAAAA:NfeRznT2-Yj06vEetmVg6pOiheTxiAdp50gvFH4H59xo3GKNUCS_NZ3Ln3iAPjpDwkyeesw_> |
|  |  |  |  |  | shap is used to figure out words that contributed to producing the given predictions. | <https://www.sciencedirect.com/science/article/pii/S0306457321000704?casa_token=y4DrAHVPOuUAAAAA:cyFCKPlER5rGmpJW_OM0CHmO3gQzpktTEbhb0ReoLA9W521xPtepVTnEN6boxRrIyfMKPdHOz90> |
| . |  |  |  |  | The coefficients from LR model were utilized to build a nomogram predicting the probability of critically ill COVID-19 related pneumonia patients | <https://www.frontiersin.org/articles/10.3389/fcimb.2022.819267/full> |
|  |  |  |  |  | (1) the DALEX package estimated the increase in the model’s binary-cross-entropy loss if each of the 708 input variables in the data set was dropped from the model one at a time. | <https://journals.sagepub.com/doi/full/10.1177/0956797620959594> |
| For data preparation, the proposed methodology employs a KNNImputer algorithm to handle null values in the dataset, isolation forest (iForest) to remove outlier data, and a synthetic minority oversampling technique (SMOTE) to balance data distribution. (3) SMOTE balances the data by randomly creating minority class instances, to over sample the minority class. |  |  | (1) The proposed model utilized structural diversity by employing two level of classifiers, the prediction from the first level classifiers (extra treess, RF, LR) was fed to the second level classifier (XGBoost) to enhance the predictor capabilities. (2) |  |  | <https://www.sciencedirect.com/science/article/pii/S2352914820305992> |
|  |  |  | (2) Recurrent Neural Networks (RNN) consider the sequence of lab test findings, instead of the snapshot of the patient’s health conditions. (3) Gradient boosting gb algorithms use a gradient descent procedure to assign a weight to each weak learner, whose value is related to the ability of the learner to reduce errors. The main differences between gb algorithms are related to the methods for building trees, error calculation, and feature selection. | (1) For selecting the best performing model, instead of evaluating the relative performance of the models on a single validation dataset, the authors follow a Bayesian pairwise comparison approach to compute the the posterior distribution, of the ‘‘worse probability’’ (i.e., the first model performs significantly worse than the second one), the ‘‘better probability’’ (i.e., the first mode performs signif icantly better than the second one), and the ‘‘Region Of Practical Equivalence’’ (ROPE), which denotes the probability of the two models having no significant differences in their performances. |  | <https://ieeexplore.ieee.org/abstract/document/10185020> |
|  |  |  | At each iteration, the genetic algorithm builds a population of policy schedules, which is fed into the deep learning model to predict the reproduction rates by incorporation time series records and country specific information. The outcome of the deep learning module is then used to parameter tuning of the SEIR model. The seir models feedback to generate better schedules, optimizing health-related objectives (e.g. minimize total deaths) while satisfying hard constraints (e.g. never exceed hospitalization capacity) is fed again into the evolutionary algorithm as its fitness measure. |  |  | <https://dl.acm.org/doi/abs/10.1145/3394486.3412863> |
| Smote resampling | (1) The number of features to utilize for model training was obtained by iteratively training an XGboost model on a collection of the top K most important features |  | Various machine learning |  | (2) a directed acyclic graphical model (Bayesian network) approach was taken to infer and visualize the effect of the potential influencers for decision making. Hill climbing optimization algorithm was used to learn bayesian networks. (4) The consensus network was then parameterized with conditional probabilities using junction tree algorithm, thus marginal probabilities and conditional probabilities were inferred using Exact Inference method of Bayesian network inference | <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0264785> |
|  |  |  |  |  |  | <https://link.springer.com/chapter/10.1007/978-3-030-72752-9_16> |
|  |  |  | The proposed ensemble learning model consists of 4 cascade levels. Each cascade level is constructed using six forests: two extra trees, two XGBoost, and two LightGBM. |  |  | <https://www.nature.com/articles/s41598-021-95957-w> |
|  |  |  |  |  | The authors applied Stable and Interpretable RUle Set (SIRUS; Bénard et al., 2021) to extract a list of rules based on the rf model. . SIRUS identified highly predictive logical rules of the form “If the abundance of feature X is greater than Y, then the probability of needing respiratory support is Z”. SIRUS inherits a predictive accuracy close to random forests, combined with the simplicity of dts. | <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1009440/full> |
|  | (1) in the feature selection phase, an initial weighted graph illustrates features with Pearson similarity criteria for the feature similarities calculation as well as integration of Fisher Score (FS) and the Node Centrality to determine the score of each feature. Hence, the feature selection approach considers not only feature importance but also feature similarity. (2) irrelevant and redundant features have presented serious challenges to the existing artificial intelligence based prediction model, impacting accuracy and prediction. Irrelevant and redundant features also increase the probability of overfitting and increase the computational complexity. |  |  |  | (3) The research belongs to the studies, which aim at transforming random forest to a single dt | <https://www.sciencedirect.com/science/article/pii/S2352914822000892> |
|  |  | The outputs of these base-learners were aggregated via a weighted voting scheme, with the weights optimized via an evolutionary paradigm. Training CNNs involves the optimization of a considerably large number of learning parameters i.e. the connection weights, filter values, etc. This optimization is done by combining the speed of gradient descent (GD) algorithms and global search space coverage of the evolutionary algorithms. | The ensemble model consists of 9 base learners including statistical approaches and convolutional neural networks. |  |  | <https://www.sciencedirect.com/science/article/pii/S2667102623000025> |
|  |  |  |  |  | a rule learner is used based on Repeated Incremental Pruning (Cohen, 1995). | <https://2wvvw.easychair.org/publications/download/V81H> |
|  |  |  |  |  |  | <https://orbi.umons.ac.be/bitstream/20.500.12907/5090/1/Paper%2001_%20Explainable%20Deep%20Learning_Covid19.pdf> |
| To identify the optimal hyper parameters of the random forest model i.e. the maximum depths of trees (‘max depth’) and maximum number of features used in each split (‘max features’), a grid search is applied. |  |  |  |  |  | <https://www.nature.com/articles/s41598-023-33033-1> |
|  |  |  | various algorithms were considered including Random forest, Gradient boosted trees, Extreme gradient boosted trees, Light-gradient boosted machine, Neural networks, Regularized regression, Elastic-net, K-Nearest neighbors, Support vector machine, Generalized additive model. multiple ensemble models are examined and searched for the model with the best accuracy. | To build the ensemble model, the average, median, elastic-net model and generalized linear model are examined and confirmed that three to five models constituted the ensemble. |  | <https://www.sciencedirect.com/science/article/pii/S2468042722000574> |
|  |  |  | (1) Time-aware Long Short-Term Memory incorporates the elapsed time information into LSTM, if diferent time intervals exist. (2) The study distinguishes 4 stages of COVID-19 progression with different patient statuses and mortality risks and |  | uses Kullback-Leibler Divergence (KL divergence) to analyze patient characteristics through each laboratory test feature. It calculates the average KL divergence of the distribution of adjacent stages. If the average KL divergence of a feature is large, it more likely is a biomarker to distinguish different stages. | <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01359-9> |
|  |  |  |  |  | Linear SVMs are also interpretable as any other linear model. Non-linear SVMs are partially interpretable, as they tell you which training data are relevant for prediction, and which aren't. This is not possible for other methods such as Random Forest | <https://link.springer.com/article/10.1186/s12890-023-02421-8> |
|  |  |  |  | An ensemble model was developed with the best-performing ML models (rf,gb,xgb,knn,svm) and a DL (MLP) model to achieve optimized results. |  | <https://www.mdpi.com/2075-4418/12/6/1464> |
|  |  | (1) The search for hyperparameters was carried out using two different algorithms; a random search and a Bayesian search. |  | (2) To evaluate the performance of the models, theerror metrics (mean absolute error—MAE, mean square error—MSE, root mean squared error—RMSE) and goodness-of-fit (R2 score) were used. |  | <https://www.mdpi.com/2073-445X/11/11/2100> |
|  | The binary SVM classifier can classify each period (r the period during the first lockdown in 2020 and a comparable non-pandemic period situation in 2019) based on the mood cluster derived by the inbeforehand k-means clustering analysis |  |  |  |  | <https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_3362723> |
| Given a dataset, we construct a set of predictors and study explanations generated e.g. by means of shap. The paper intoduces the issue of underspecification when observing “too many” different explanations form various predictors on a dataset. It suggests to use shap as it is applicable to any data type and ml method. the authors show how e.g. by increasing the size of data, explainations of the random forests converge towards each other. |  |  |  |  |  | <https://link.springer.com/chapter/10.1007/978-3-030-89188-6_24> |
|  | The paper uses a form of Latent Dirichlet Allocation, a Bayesian unsupervised clustering method that can naturally model the categorical nature and inherent imprecision of disease codes. This clustering distils thousands of disease codes, and billions of possible combinations into a set of ‘digital comorbidity fingerprints DCFs. Each patient’s pre-existing conditions are thus summarised by a single number (between 0-100%) describing how strong a specific pre-existing condition fingerprint is present in their record. All pre-existing conditions and their myriad of combinations are summarised by a total of 30 DCFs. DCFs offer immediately interpretable clinical descriptions that are meaningful |  |  |  |  | <https://www.medrxiv.org/content/10.1101/2021.03.29.21254579v1> |
|  |  |  | (1) the paper proposes a hybrid ml-dl model: first, by adding a 1\*10 fully connected layer before the final output layer, a modified dcnn (resnet) is created to transform the imaging data to a 10 features output model, then the clinical information (10 feutures) and lab testing information (23 features) of patients are added to the 10 extracted image features. RF, SVM, and kNN models are utilzed to classify the severity. (2) |  | gini impurity score is computed based on random forest model. (2) | <https://www.jmir.org/2021/1/e25535/> |
|  |  |  |  |  | cam exploits global average pooling and allows to compute class-specific heatmaps that indicate the discriminative regions of the image that caused the particular class activity of interest | <https://towardsdatascience.com/ultrasound-for-covid-19-a-deep-learning-approach-f7906002892a> |
|  |  |  |  |  |  | <https://www.mdpi.com/2227-9032/9/9/1099> |
|  |  |  |  |  |  | <https://www.nature.com/articles/s41746-021-00453-0> |
|  | (1) First, the paper uses HHOSRL as the optimization algorithm to find the optimal subset (3) HHO is a new swarm algorithm that was proposed inside the class of bio-inspired methods, used for feature selection. |  | , and after finding the optimal features, it uses KELM as the classifier for the classification task. (2) KELM is a special kind of single-implicit feedforward neural network |  |  | <https://www.sciencedirect.com/science/article/pii/S0010482521009604?casa_token=xDVkrDV2jgUAAAAA:iplqPXRkLKGCe6AupBiU6RUkWOMv3wrcbjQxSAsvVKGTbznhKVfpJ5AE-K5c-yTZfqUwKNmZriQ> |
|  |  |  | (1) XGBoost has built-in capabilities to handle missingness, sparsity and feature interactions |  |  | <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0285991> |
|  |  |  |  |  | The human attention visual mechanism uses limited attention to quickly screen high-value information from a large amount of information. The attention mechanism in deep learning is essentially similar to the selective visual attention mechanism of humans. Its core goal is to select information that is more critical to the current task goal from a large number of information. At present, the attention mechanism network has been widely used in various deep learning applications such as natural language processing, image recognition, and speech recognition. In this article, | <https://www.researchsquare.com/article/rs-577494/v1> |
|  |  |  |  |  | rf feature importance reports the importance of each individual variable to the model by quantifying the percentage of trees that use the variable. | <https://www.acpjournals.org/doi/full/10.7326/M20-6754> |
|  | The feature extraction role is performed by a deep, pre-trained specialized CNN, but the regression and classification results are given by a simple logistic regression model. |  |  |  |  | <https://arxiv.org/abs/2011.14983> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9514609> |
|  |  |  |  |  | The study calculated the average marginal effects using the idea of calculating partial dependence (Molnar, 2020) | <https://www.sciencedirect.com/science/article/pii/S2214367X23000352> |
|  |  |  |  |  |  | <https://www.mdpi.com/1999-4915/14/3/625> |
|  | (1) pca is used for dimensional reduction, lasso for feature selection. the PCA algorithm projects the features to a new dimension, and the features on the new dimension become mixed features. These new features cannot provide a strong explanatory basis |  |  |  |  | <https://dl.acm.org/doi/abs/10.1145/3543377.3543384?casa_token=87bZtcn6Dc4AAAAA:yZBvULd3uHXsqLHy4tTrE6o2npmYqUWiH8Kr9BX37bNQHsf-apsUw4E8y9a3nLg0QaO8IAcelgknMw> |
|  |  |  |  |  |  | <https://www.frontiersin.org/articles/10.3389/fpubh.2022.987376/full> |
|  |  |  |  |  |  | <https://www.nature.com/articles/s41598-022-18472-6> |
|  |  |  | Gradient Boosted regression trees and Support Vector Regression are used (not classification). |  |  | <https://www.ceeol.com/search/article-detail?id=1111407> |
|  |  |  |  |  |  | <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00476-0> |
|  |  |  |  |  |  | <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3590467> |
|  |  |  |  |  | (1)The derivation of relevance scores value goes from upper-layer neurons to lower layer neurons | <https://www.frontiersin.org/articles/10.3389/fimmu.2023.974343/full> |
|  | The paper introduces a non-supervised clustering analysis with neural network self-organizing maps (SOM). The outcome of the analysis process is the allocation of each sample in a cluster by maximizing the similarity between elements in the same cluster. The SOM used here grouped 67 of the 81 positive SARS-CoV-2 patients into three principal map units (15, 20, 25) that had common patterns and behaviors. Thus, a significant portion of the negative results (246 from 518) was separated.. |  |  |  | during the process of finding and establishing the “Information Need,” which in this case was determined as “what variables in routine laboratory blood test at present significant patterns for detection of SARS-CoV-2 positive patients,” the SOM analyses demonstrated some attributes weights (that represented the variables in the blood test) in each unit of the map, generating human explainability and interpretability of the decision performed in the training process. The main variables found in SOM experiments were: Leukocytes, Basophils, Eosinophils, and Red Blood Cell Distribution Width | <https://link.springer.com/article/10.1007/s00500-021-05810-5> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9881792?casa_token=tqOwRWJGnv4AAAAA:orRTZhj1TkWQ88she2JdpCGIlM8QxXjcPL2pYQnPEcrLlgPhU4xK3irpfNGcC1MTLchpmLy2mLHHAw> |
|  | The feature selection process consisted of three steps. Firstly, the features were ranked by importance and frequency scores computed from the na¨ıve XGB model. Then, the importance and frequency scores were clustered by the k-means clustering method. Finally, the features were selected based on the significance of the cluster at 99%. |  |  |  |  | <https://downloads.hindawi.com/journals/jat/2022/6458371.pdf> |
|  |  | Hyperparameters were optimized for all steps in the pipeline with 150 iterations of Bayesian optimization |  |  |  | <https://www.jacc.org/doi/abs/10.1016/j.jacadv.2022.100043> |
| (1) SMOTE is used to create synthetic data that is close, or nearest neighbor, to the minority class in the feature space | (2) pca is performed to compute the minimal set of principal components that explained 95% of the variance in the data. recursive feature elimination rfe is used to to select the minimal set of predictors that were most predictive for a COVID-19 positive test. |  |  |  |  | <https://link.springer.com/article/10.1007/s11517-022-02549-5> |
|  |  | 4 types of loss functions for dnn keras model are examined to figure out the best accuracy |  |  |  | <https://revistas.unal.edu.co/index.php/dyna/article/view/105616> |
|  | K-Means is one of the popular clustering algorithms that can detect complex patterns. This algorithm works with a partition system to group data into several clusters. |  |  |  |  | <https://journal.universitasbumigora.ac.id/index.php/Varian/article/view/2149> |
|  | t-stochastic neighborhood embedding t-SNE is an unsupervised clustering dimensionality reduction technique. The 2D map allows for the simultaneous assessment of the relative similarity of all subjects in the dataset, along with the distribution of their clinical characteristics. Axes represents arbitrary unit determined based on t-SNE. Target variables are indicated by color-bar in each mapping. |  |  |  |  | <https://www.nature.com/articles/s41598-022-17567-4> |
|  | To reduce the embedding dimension to the 5-dimension clustering techniques e.g. UMAP is used. The results is presented by Scatter plot of UMAP embedding of the topics related to the first, second and third waves of the pandemic. |  | BERT is pre-trained language model based on Transformers, i.e. encoder-decoder LSTM-based recurrent neural network that features enhanced memory mechanism known as attention. The encoder generates an encoding that feature information about the relevant parts of the inputs, which is passed to the next encoder layer as inputs. The decoder layer does the opposite of the encoder to generate an output sequence. Transformer models implement the mechanism of attention (by weighting the significance of each part of the input data, which has made them prominent for language modelling tasks). Although BERT is pre-trained, it is usually trained further with datasets for specific applications. BERT on its own cannot be used for topic modelling, it only provides a word embedding that would be an input for clustering methods. |  |  | <https://arxiv.org/abs/2303.00135> |
|  | BorutaShap algorithm, A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values, is applied to select feuatures. Based on the selected features, as set ofa algorithms especially a random forest granting the best performance is applied. |  |  |  |  | <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9511418/> |
|  | Countries devided into 4 clusters without conveying interpretation of this result |  |  |  |  | <https://dergipark.org.tr/en/pub/eams/issue/72229/1084717> |
|  |  |  | The paper leverages the temporal character of COVID-infections by incorporating the NPIs and the number of cases and deaths. the input to the model is a stack of sequences. Each sequence stands for a different variable. Every element in a sequence shows the daily value of the variable covering a time period of a week, thus in each sequence there are seven elements. The model predicts two sequences i.e. daily cases and deaths over the next week. it applies 4 types of rnns. However interpretation of results is missing. |  |  | <https://www.mdpi.com/1424-8220/22/10/3658> |
|  | ISOMAP , Isometric Mapping was used to generate a compact representational form of raw features. The optimized feature representation was thereafter used to train the GBR framework. |  |  |  |  | <https://www.mdpi.com/2227-9091/11/5/94> |
|  |  |  |  |  |  | <https://www.frontiersin.org/articles/10.3389/fpubh.2021.779501/full> |
|  |  | Two pre-trained neural networks, ResNet50 and MobileNet, were fine-tuned for the datasets under analysis. |  |  |  | <https://link.springer.com/chapter/10.1007/978-3-031-30788-1_7> |
|  |  |  |  |  |  | <https://link.springer.com/chapter/10.1007/978-3-031-04597-4_1> |
|  |  |  |  |  |  | <https://www.medrxiv.org/content/10.1101/2021.07.07.21260097v1> |
|  | This study investigates the efficacy of three feature extraction techniques, including Mel Frequency Cepstral Coefficients (MFCC), Chroma, and Spectral Contrast features |  |  |  |  | <https://arxiv.org/pdf/2309.04505.pdf> |
| A data augmentation procedure based on autoencoders (AEs) was proposed to balance the dataset. Details of the implementation of these methods are available in MEALPY61, which is a Python module consisting of meta heuristic algorithms. regardless of the feature selection method, the CNN-AE trained on the selected features did not outperform the CNN-AE trained on the full dataset. This is because the CNN already included an automatic feature selection mechanism and could rule out unnecessary features during learning. | AEs belong to the realm of unsupervised learning, as they do not need labelled data for their training. In brief, an AE compresses input data to a lower dimensional latent space and then reconstructs the data by decompressing the latent space representation. Similar to principle component analysis (PCA), AEs perform dimensionality reduction in the compression phase. However, unlike PCA, which relies on linear transformation, AEs carry out nonlinear transformation using deep neural networks. The authors also examine whether feature selection improves the classifcation performance of the clinical dataset. We relied on meta-heuristic population-based algorithms to carry out feature selection. Te meta-heuristic methods that have been used in the experiments are Artifcial Bee Colony (ABC)55, Ant Colony Optimisation (ACO)56, Butterfy Optimisation Algorithm (BOA)57, Elephant Herding Optimisation (EHO)58, Genetic Algorithm (GA)59 and Particle Swarm Optimisation (PSO)60. |  |  |  |  | <https://www.nature.com/articles/s41598-021-93543-8> |
|  | Spatial lag regression models (SLR-models) were calculated for the top shap-based 20 features leading to simillar associations to the shap values |  |  |  |  | <https://www.mdpi.com/1660-4601/18/20/10663> |
|  |  |  |  |  | lasso logistic regression yields predictive performance for survival with mechanical ventilation similar to or better than the alternatives while having the benefit of being interpretable. For duration of mechanical ventilation, we find the hurdle models to produce estimates with a mean absolute error (MAE) value that is lower than those of linear regression and regression trees, and comparable to the MAE of xgBoost. Again, as transparency and interpretability are important, we select the hurdle model over the xgBoost approach. | <https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.13934?casa_token=ASnQcokdcoEAAAAA:7nUzq2s4Mp4a_lF34J8V1V8S9oXRyDM9MdaN-n8qI86HKAKwxjcLiw2ycyKg4VCx6vvHD_x1fOI9u5o> |
| Before GBM modeling, missing data were imputed using the missForest imputation algorithm. |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S2542454822000455> |
| (The authors investigate whether the inclusion of day specific effects – i.e., 366 time dummies ( that e.g. capture unobserved variations such as seasonal changes, synchronized fears of infection, etc.) and corridor-specific effects – i.e., 45 corridor dummies (that e.g. capture unobserved variations such as the skill level of the cross-border work force, linguistic and cultural proximity between countries, etc.) improves the predictive power of the models. The inclusion of 45 corridor specific dummies always improves the quality of fit. On the contrary, the inclusion of 366 day-specific dummies deteriorates the performance of KNN and G-Boost methods. the gains from adding information about unobserved common time trends, is indeed outbalanced by the costs linked to the inflated dimensionality of the computation problem. | 1)The authors use a PCA analysis to reduce the dimensionality of the origin- and destination specific containment measures and extract the first two components of the PCA and propose the first PCA component can be interpreted as an average index of stringency of containment measures and the second component captures testing and tracing policies. (2) ML techniques always outperform the linear EM model. (3) |  |  |  |  | <https://globalizationandhealth.biomedcentral.com/articles/10.1186/s12992-022-00832-6> |
| data set is small size and imbalanced |  |  |  |  |  | <https://arxiv.org/abs/2209.10664> |
|  |  |  |  |  | (1) BERT and its variations split the texts into pieces of words called tokens instead of words they receive as inputs for the model. (2) BERT defines and adjusts attention weights in each layer and head, giving higher weight to the tokens that it considers most important.(3) The paper proposes the concept of word attention, in which attention weights assigned to tokens are transferred back to the original words in each tweet. | <https://sol.sbc.org.br/index.php/sbbd/article/view/17867> |
|  | differently from complex, black-box ML methods such as random forest, regression trees allow an intuitive understanding of the mechanism through which the outcome variable of interest is linked to its most relevant predictors |  |  |  |  | <https://mpra.ub.uni-muenchen.de/104404/> |
|  |  |  |  |  |  | <https://www.mdpi.com/2075-4426/12/8/1325> |
|  | dt is used for feature selection |  |  |  |  | <https://library-archives.canada.ca/eng/services/services-libraries/theses/Pages/item.aspx?idNumber=1369198255> |
|  |  | proposed framework for COVID-19 forecasting is an extension to the standard Susceptible-Exposed-Infectious-Removed (SEIR) model. The model introduces in addition compartments for hospital resource usage as they are crucial to forecasts for COVID-19 healthcare planning. Learnable encoders infer the transition rates at which individuals move through different compartments, trained on static and time-varying public data, to model the changing disease dynamics over time. The paper adopts a generalized additive model for each variable based on features at different times t-1 to t-k. |  |  |  | <https://www.nature.com/articles/s41746-021-00511-7> |
|  |  |  |  |  |  | <https://www.tandfonline.com/doi/abs/10.1080/10584609.2023.2201184> |
|  | PCC assesses the linear relationship between two variables and provides the correlation degree within a range between 1 and −1. |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0952197622006078> |
|  | A chi-square test was performed to evaluate the association between independent variables and the dependent variables |  |  |  |  | <https://www.mdpi.com/2227-9032/11/14/2080> |
|  |  |  |  |  |  | <https://www.mdpi.com/2254-9625/13/9/120> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/10159467> |
| Univariate logistic regression was used to select potential explanatory variables for the outcomes. For multivariable analysis, variables with too many missing values (.5%) were removed. Variables that were significant based on univariate analysis were placed in the order of clinical importance (demographics, comorbid conditions, severity of illness, medications, and other measures), and then backward stepwise logistic regression was used to select the best model. |  |  |  |  |  | <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8967640/> |
| The histogram of the LASSO logistic regression weights represents the importance of each feature for the classification task. The positive weights reflect a susceptible behaviour of the features to the target COVID-19 disease, whereas the negative weights a protective action. |  |  |  |  |  | <https://www.medrxiv.org/content/10.1101/2020.11.04.20225680.abstract> |
| Authors also examine different BN structure learning algorithms to examine how these perform in predicting the risk of COVID-19. |  |  | (1)In the past, ML approaches have been taken to work with structured EHR data. Simpler methods often limited the time information and just worked with a one-hot encoding (OHE) of diagnoses and prescriptions, which allowed the application of standard ML techniques, such as logistic regression, random forest (RF), XGBoost (XGB), and Bayesian methods. Recently, more studies focused on the use of time-series information. Methods for such an approach include autoencoders, convolutional neural networks, or sequential models like recurrent neural networks (RNN) and transformer-based models. Transformer-based models originate from natural language processing (NLP) and have recently gained much attention. A principal advantage of transformer models is the ability to train them in a parallel fashion and to weigh different parts of a time series differently due to their inbuilt attention mechanism. Transformer-based models typically undergo two-stage training: pre-training for generic representation learning and transfer learning (fine-tuning) for application specific prediction. This approach enables sharing pre-trained models, often based on large datasets like the entire Wikipedia or protein sequences, with a broader community. These models can then be fine-tuned for various unforeseen tasks, highlighting the transformer-based approach’s versatility and strength. |  |  | <https://link.springer.com/article/10.1007/s41324-021-00421-6> |
|  |  | Due to the high computational costs for information processing and finding the optimal CNN parameters (number of layers, number of filters per layer, filter size and batch size), a new swarm (evolutionary) intelligent based meta-heuristic optimization algorithm is proposed. |  |  |  | <https://www.researchgate.net/profile/Odeniyi-Olufemi-Ayodeji/publication/371339460_Development_of_a_COVID-19_Patients%27_Fatality_Prediction_System_Using_Swarm_Intelligent_Convolution_Neural_Network/links/64ff58ee849bbb203b913409/Development-of-a-COVID-19-Patients-Fatality-Prediction-System-Using-Swarm-Intelligent-Convolution-Neural-Network.pdf> |
| Given the lack of high quality labeled data, semi-supervised methods are implemented to perform COVID-19 detection, making use of cheaper unlabeled data to improve the model’s accuracy. |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9445026/> |
|  | (1) selected figures are performed by adaboost |  |  |  |  | <https://www.mdpi.com/2076-3417/11/21/9865> |
|  |  |  | The paper proposes a Gated Attention based CNN, which gets the input data in a 1-Dimensional form of length of clinical blood parameters |  |  | <https://ieeexplore.ieee.org/abstract/document/10040102/?casa_token=0TJbKj1J_Z0AAAAA:KJ4Dl7kDcA-AJVk5mgIBINkmIezTjGw5MhmMBiwIylazpoT6RbqvhfGgaArF1H_B3OZBfR7gfimFLA> |
|  |  |  |  |  |  | <https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(22)00047-5/fulltext> |
|  |  |  |  |  |  | <https://www.frontiersin.org/articles/10.3389/fmicb.2021.729455/full> |
|  |  |  | (1)Ordinary Least Squares (OLS) vs Kernel-based Regularized Least Squares (KRLS) estimations: The KRLS estimator can learn the data to decrease potential misspecification bias. The KRLS method can also solve potential problems of non-linearity and multicollinearity in the OLS estimations. (2) In all the KRLS estimations, the only statistically significant coefficients are internal and external conflicts. |  |  | <https://www.frontiersin.org/articles/10.3389/fpubh.2021.681604/full> |
|  | (1)For the sake of exploratory analysis, a hierarchical clustering using a bottom-up approach was applied to the correlation between scales. This method performs a hierarchical clustering in which each scale starts in its cluster, and clusters are successively merged. The metric used for the merge strategy was the complete linkage, which minimizes the maximum distance between observations of pairs of clusters. (2) | CatBoost has performance similar to other algorithms such as LightGBM and XGBoost. However, one of the main advantages of CatBoost is that it achieves good performance using default hyperparameters, with no need to tune them. |  |  |  | <https://www.science.org/doi/full/10.1126/sciadv.abj7205> |
|  |  |  |  |  |  | <https://www.tandfonline.com/doi/abs/10.1080/2573234X.2022.2155257> |
|  |  |  |  |  |  | <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0293147> |
|  |  |  | The paper proposes the policy-aware temporal convolutional network (P-TCN) model that considers the policy dependence of time-series data irregularities. The model employed 3D vector data (consisting of policy attribute features, time steps, and residential parking occupancy-related features) as input to predict future parking occupancy. |  |  | <https://www.sciencedirect.com/science/article/abs/pii/S0965856423002525?casa_token=rvJN5VqjZmcAAAAA:qjzU88ixkPOcJbVNjQAueFBOXk4cRAbM7zItmg-yKZOlQiXTcU1Q80AyGljW-S7-IoJAxHW1HNU> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0966692323000662> |
| The seasonal ARIMA model contains both non-seasonal and seasonal components. |  |  |  |  |  | <https://www.atlantis-press.com/proceedings/msea-22/125982675> |
|  |  |  | KRLS is an appropriate method to use when the assumptions of linear regression are not met and the precise functional forms between the predictors and outcomes are unknown. |  |  | <https://repository.up.ac.za/handle/2263/84660> |
|  |  |  |  |  |  | <https://royalsocietypublishing.org/doi/full/10.1098/rsif.2021.0702> |
|  |  |  |  |  |  | <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4102003> |
|  | Conventional CNNs perform convolution operations in the lower layers of the network and concatenate the last convolution layer’s feature map to the fully connected layer, followed by a softmax logistic regression layer, for classification. |  |  |  |  | <https://www.frontiersin.org/articles/10.3389/fcimb.2023.1116285/full?utm_source=dlvr.it&utm_medium=twitter> |
|  |  |  |  |  |  | <https://erj.ersjournals.com/content/59/2/2100284.short> |
|  |  |  |  |  |  | <https://pure.ulster.ac.uk/en/publications/house-prices-airport-location-proximity-and-air-traffic-volume-pl> |
|  |  |  |  |  |  | <https://www.frontiersin.org/articles/10.3389/fmars.2022.911819/full> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0925753523001935> |
|  |  |  |  |  |  | <https://www.medrxiv.org/content/10.1101/2020.08.29.20184366v2> |
|  |  |  |  |  |  | <https://ascelibrary.org/doi/abs/10.1061/JWRMD5.WRENG-5650> |
|  |  |  |  |  |  | <https://www.medrxiv.org/content/10.1101/2020.04.27.20081562v1> |
| (2) data unbalance: During the fitting procedure, the class slight unbalancing is tackled by penalizing the misclassification of the minority class with a multiplicative factor inversely proportional to the class frequencies. (3) Penalized classification imposes an additional cost on the model for making classification mistakes on the minority class during training. |  | (1)hyper parameter tuning: is tuned with a grid search method on the average area under the ROC curve for the 10-fold cross-validation. |  |  |  | <https://www.tandfonline.com/doi/full/10.1080/15548627.2021.1995152> |
|  |  |  |  |  |  | <https://www.mdpi.com/1660-4601/18/18/9629> |
|  |  |  | (1)GAMs are used as nonlinear regression tools that allow for nonparametric fittings of complex dependences of response |  |  | <https://online.ucpress.edu/elementa/article/10/1/00044/169476/Photochemical-sensitivity-to-emissions-and-local> |
|  |  |  |  |  |  | <https://onlinelibrary.wiley.com/doi/full/10.1111/ssqu.13295> |
|  |  |  |  |  |  | <https://www.researchgate.net/profile/Juracy-Bertoldo-Santos-Junior/publication/353231152_Inteligencia_artificial_e_COVID-19/links/61f2b8bfdafcdb25fd55cc5f/Inteligencia-artificial-e-COVID-19.pdf> |
|  |  |  |  |  |  | <https://www.proquest.com/docview/2839529574?pq-origsite=gscholar&fromopenview=true> |
|  |  |  |  |  |  | <https://repositorium.hs-ruhrwest.de/frontdoor/index/index/docId/782> |
| A completely different imputation approach is used by Generative Adversarial Neural Networks (GANN), which learn to generate ‘‘missing’’ data with the same distribution as the training set. This is done by training a ‘‘generative’’ network, which generates possible imputed values and proposes them to a ‘‘discriminative’’ network, which is trained to accept only those generated values that properly fill the missing ones according to the underlying data distribution. |  |  |  |  | (1) even if ATs (associative dt) can provide simple and human understandable decision rules, a limitation of this approach is that the resulting model does not exactly fit the original RFs(2) | <https://ieeexplore.ieee.org/abstract/document/9239931> |
|  |  |  | This study uses CNN-LSTM, a hybrid deep learning technique. The first layer is the CNN layer; the data will then pass through the convolution layer, where the filters will extract the most critical features to generate a feature map. This map will undergo max pooling to preserve the most dominant features, followed by batch normalization. The output will be sent to an LSTM layer to extract temporal features. |  |  | <https://www.sciencedirect.com/science/article/pii/S2666307423000037> |
|  | The paper used the SHAP- Recursive Feature Elimination SHAP-RFECV from the Probatus python library to perform the feature selection and shap for feature importance. |  |  |  |  | <https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2021.662343/full> |
|  | (1) 6 clustering methods are used to identify the groups of subjects with amelioration, deterioration and stability to their mood state: K-Means, Spectral Clustering, Ward, Average Linkage, Balanced Iterative Reducing and Clustering using Hierarchies (Birch), and Jenks natural breaks optimization method (Jenks) (2) The study of different application of feature selection methods e.g. Chi-Square, ANOVA, MI, CMIM, MRMR, SURF\*, MultiSURF, MultiSURF\*, and ReliefF demands pusuing indepth research in the related fields. |  |  |  |  | <https://www.mdpi.com/2227-9032/10/1/149> |
| (1)SMOTE deploys principles of interpolation (3) (5). |  | (2) Hyperparameter tuning was conducted for all the ML classifier pipelines with GridSearchCV. GridSearchCV tests all combinations of hyperparameters and narrows the model parameters to the optimal ones. |  |  | Quantum Lattice: This tool is inspired by the Richard Feynman’s path. This is a supervised machine learning mehtod to analyze the data and accordingly creates multiple possible graphical models composed of different mathematical operations (opposed to black box models). On selection of the best model, a Qgraph is created to provide the rationale behind a prediction. Further, a simplified equation for the model is obtained that provides insights into the mapping of inputs to outputs.(6)Anchors are local High-model-agnostic explanations of features that assist provide high precision predictions explained by decision rules | <https://www.sciencedirect.com/science/article/pii/S2772662223000863> |
|  |  |  |  |  | gini impurity is used to reduce the dimensionality of the features. The mean decrease in Gini coefficient is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting random forest | <https://link.springer.com/article/10.1007/s10844-022-00768-8> |
|  |  |  |  |  | (1) training set-based gini impurity importance doesn’t reflect the actual performance of the feature on the test set. To avoid such bias, Breiman proposed Permutation based feature importance pfi (2) pfi and shap and all ml methods are provided by the sci-kit learn module in Python | <https://www.preprints.org/manuscript/202206.0115/v1> |
|  |  |  |  |  | (1) Post-hoc systems first treat ML models as a black-box and give specific explanation for instances after the model training. Ante-hoc systems, conversely, embed explainability into the model during the training time, and thus build an explanation model using a white-box (or open-box) approach. LIME will tune the values of the features from the picked instance and generate new samples based on the proximity to the instance being picked. At last, LIME will optimize a line based on all generated samples and give a local interpretable explanation of the instance being picked. Gini Importance presents the total decrease in the node impurity averaged over all trees of the ensemble model. | <https://arxiv.org/abs/2105.13430> |
|  |  |  | The stacking model involves multiple base models and a meta logistic regression model that combines the predictions of the base models. Unlike bagging, in stacking, the models are typically different (e.g. not all decision trees) and fit on the same dataset (e.g. instead of samples of the training dataset). Unlike boosting, in stacking, a single model is used to learn how to best combine the predictions from the contributing models (e.g. instead of a sequence of models that correct the predictions of prior models). (2) the authors use also a deep covolutional resnet. Residual Network (ResNet) is a Convolutional Neural Network (CNN) architecture that overcame the “vanishing gradient” problem, making it possible to construct networks with up to thousands of convolutional layers, which outperform shallower networks. (3) In practice, Convolutional Networks like resnet get pretrained on very large datasets like image net, and then are used either as an initialization or as a fixed feature extractor for the task of interest. (4)However, the pertained models they may not be fully compatible or transferable to your task. In addition, they may be difficult to interpret and understand, and may contain biases and errors.(5)CNNs can still achieve comparable performance when data is scarce. This trait still makes them relevant for various computer vision projects. https://arxiv.org/pdf/2010.11929.pdf |  |  | <https://www.medrxiv.org/content/10.1101/2021.11.01.21265527v1> |
|  |  |  | (1)The stacking model involves multiple base models and a meta learner. |  |  | <https://www.tandfonline.com/doi/full/10.1080/23311916.2023.2272361> |
|  | The pipeline comprises a text-guided visual feature generator to generate comprehensive visual features from the news text, as well as an image-guided textual feature decoder to generate the corresponding textual features from the news image. |  |  |  |  | <https://dl.acm.org/doi/abs/10.1145/3485447.3512257?casa_token=23L8mrvDQDoAAAAA:-eIs_g3DVY6tdqD4DOcRfHlv4_K0xQ0vlnxQRjquwE2Sp8UY_Bg1Ex7_Bmniyv1lnQRvtjIcH2iDwQ> |
|  |  |  |  |  |  | <https://www.nature.com/articles/s42003-022-04073-6> |
|  |  |  | The paper uses Bayesian geospatial modeling. Poisson regression, Poison mixed effect model, Poisson Bersag-York-Mollie (BYM) spatial model, Negative-Binomial (NB) regression, NB mixed effect model, NB BYM spatial model are utilized as statistical Bayesian models. Bayesian frameworks can consider flexible uncertainties in model parameters and incorporate prior expert knowledge with hierarchical structures. |  |  | <https://www.nature.com/articles/s41370-023-00518-0> |
|  | Countries with similarly balanced diets were clustered using the SOM method. In addition to the som Self-organizing map clusters for clustering countries due to the dietary data set, a Decision map corresponded to the SOM clusters presented with squares representing the most dominant feature leading to the decision for clustering the countries. |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0957417422014907> |
|  |  |  |  |  | The paper uses an adopted version of the PGM explainer (to work on our tabular dataset), which is an explainer for graph neural networks gnns. the explanations of a GNN’s prediction is a simpler interpretable Bayesian network approximating that prediction through probabilistic understanding of the effect of perturbing the given features. Since Bayesian networks do not rely on the linear-independence assumption of explained features, PGM-Explainer is able to illustrate the dependency among explained features and provide deeper explanations for GNNs’ predictions than those of linear or additive feature attribution methods. | <https://arxiv.org/abs/2304.06953> |
|  | The main purpose of SOMs is to cluster the high-dimensional data into easy-to-visualize outputs. SOM is different from typical “Artificial Neural Networks (ANNs)” in its architecture and algorithmic properties. The organizing maps use competitive learning whereas the ANNs use error correction learning such as back-propagation with gradient descent. |  |  |  |  | <https://www.worldscientific.com/doi/abs/10.1142/S0218348X22401223> |
|  |  |  |  |  |  | <https://dl.acm.org/doi/abs/10.1145/3477495.3531745?casa_token=UeAsa2ZUNSQAAAAA:cc5l-qtouLQK4RYyEl3a29xhG8bWLliPC4kBayERzjNHU75eEamULw4LhW9gMJAGIUWZlA3Lu8FVFg> |
|  | Principal Component Analysis (PCA) was used on features extracted, in order to reduce the dimensionality from 40 to 5 features. Features extracted, alongside with laboratory and clinical data, are fed for selection in a Boruta algorithm with SHAP game theoretical values. |  |  |  | interpretability at global level and especially as single prediction level is given by the SHAP analysis. | <https://www.nature.com/articles/s41598-022-07890-1> |
|  |  |  |  |  | Class activation map (CAM) uses the notion of global average pooling (GAP) and learns weights from the output of the GAP layer onto the output classes. Grad-CAM generates a localization map that shows the critical places in the image representing the lesions by using gradients from the target label/class settling into the final convolutional layer. Grad-CAM++ generates a pictorial depiction for the class label as weights derived from the feature map of the CNN layer by considering its positive partial derivatives. Score-CAM does not incorporate the use of gradients like grad-cam , as the propagated gradients introduce noise and are unstable. Score-CAM is built on the idea of perturbation-based approaches that mask portions of the original input and measure the change in target score. The produced activation mask is handled as a mask for the input image, masking sections of the input image and causing the model to predict the partially masked image. The target class score is then used to reflect the significance of the class activation map. | <https://www.mdpi.com/2075-4418/12/6/1482> |
|  |  |  |  |  |  | <https://www.mdpi.com/2379-139X/8/4/151> |
|  | feature selection comprises removing correlated variables to avoid collinearity (with pairwise correlation coefficient greater than 0.8), removing low influence variables by seperating the observations according to the two targeted classes and performing a two-sided t-test retaining the features for which we have a 95% confidence that the mean for the two samples is different and a sebsequent Cross-Validated Recursive Feature Elimination comprising iteratively elimination of the least important features (e.g. based on coeffcients of the regression or shap). |  |  |  |  | <https://www.nature.com/articles/s41598-021-95004-8> |
|  | forward feature selection and backward feature elimination is performed by adding features one by one to the feature set according to their shap feature importance |  |  |  |  | <https://www.mdpi.com/2076-393X/10/10/1747> |
|  |  | hyper parameters are often searched by grid search |  |  |  | <https://dl.acm.org/doi/abs/10.1145/3524458.3547236> |
| missing values are inherently handled by the gradient-boosting predictors. |  |  |  |  |  | <https://www.nature.com/articles/s41746-020-00372-6> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0010482522008733?casa_token=NCFKMSUSicoAAAAA:sDwW-Jfka1IacWJOKpkvKqAk5QRWBCBQfAMJWBggkkfPT_qK5ITBPDkcfZZdPttIRwa2lUC_lco> |
|  |  |  |  |  |  | <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9285142/> |
| . | Word Embedding is a feature learning technique where words are mapped to vectors of real numbers |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S2212420921000674?casa_token=ZwwESdzfejAAAAAA:oD5V7jNNzguI_CAP88lQQH63CBbbMlaGTy7y3KYUZPhXOhN7S0wM6MvWUj4dtMSnhtBCNWCTdUI> |
|  |  | Hyper parameter tuning: The learning rate achieved several values ranging from 0.00001 to 0.01. The results showed that Bert Tokenizer LSTM model responds better with a very low learning rate; specifically, the learning rate in this case was equal to 0.00001. On the other hand, simple LSTM has a learning rate of 0.005, and the other models perform better with the default Keras value, that is, 0.001. |  |  |  | <https://link.springer.com/article/10.1007/s00521-022-07650-2> |
|  |  | Hyper parameter tuning and evaluation: The hyperparameters of XGBoost were optimized using Bayesian optimization, with tenfold cross-validation in the training set to overfit the model |  |  |  | <https://link.springer.com/article/10.1186/s12879-022-07317-2> |
| (2)As the amount of data increases, the epistemic uncertainty related to the model decreases. |  |  | (1)The approach of the paper to icorporate uncertainty in the traning is to perform an ensemble of different pertained cnn models. |  |  | <https://www.mdpi.com/2075-4418/13/3/441> |
| Conditional adversarial network gan is used for synthetic data generation of covid and non-covid ct images |  |  |  |  |  | <https://link.springer.com/article/10.1007/s12559-021-09926-6> |
|  |  |  | The inverted bell curve is used to assign weights to the classifiers’ outputs. The more we move further from the centre of the bell we attain higher weight values, and thus the shape of the inverted bell is utilized to calculate the weight for an output vector. |  |  | <https://link.springer.com/article/10.1007/s00521-021-06737-6> |
|  | (1)To check if there is any natural clustering in Clinical Data, the paper uses the t-distributed stochastic neighbor embedding (t-SNE) approach.(2) (4) they use an unsupervised feature selection which is an approximate kernel method called Random Fourier Features (RFF). The RFF maps the given data to a low dimensional randomized feature space (euclidean inner product space). |  |  |  | in order to compute the importance of each attribute with respect to class labels, the entropy based concept of the Information Gain (IG) between each attribute and the true class label are computed. | <https://link.springer.com/article/10.1007/s11517-022-02570-8> |
| to harness as much information as possible from the limited available information at hand, an unsupervised learning methodology for compression, feature extraction, and denoising at early stages of the network (using AutoEncoders) can result in a tailored task-specific understanding of the data (\*) |  |  |  |  | Perturbation-based explanation methods often measure the contribution of an input feature to an image classifier's outputs by heuristically removing it via e.g. blurring, adding noise, or graying out, which often produce unrealistic, out-of-samples. Instead, the authors propose to integrate a generative inpainter into attribution methods to remove an input feature. | <https://www.medrxiv.org/content/10.1101/2020.04.14.20065722v1> |
|  | (1)Feature selection, model training and interpretation are interconnected. BorutaShap is an efficient Python-based wrapper method that merges the Boruta feature selection algorithm with SHAP. BorutaShap is only compatible with and supports tree-based learners as the base model. In the process of identifying the most influential features, the Boruta algorithm generates shadow features, exact replicas of the original features, and systematically randomizes their values to eliminate any correlations with the response variable. These original and shadow-shuffled features are then employed in the tree based model to predict the target variable, leveraging the strength of the respective learner. Consequently the algorithm calculates the permutation feature importance for both the actual features and the shadow-shuffled inputs for all trees. Only when the importance score of the actual feature surpasses a certain threshold is the feature regarded as significant. Moreover, the algorithm conducts comparative analyses between the features and their corresponding shadow features, evaluating their Shapley importance values (SHAP values), which yields a more consistent outcome. |  |  |  |  | <https://www.mdpi.com/2227-7390/11/14/3145> |
|  |  |  | (4) The Reptree machine learning algorithm is of the supervised tree family and is deployed for the purpose of rule generation. |  |  | <https://www.sciencedirect.com/science/article/pii/B9780128245361000046> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/abs/pii/S2173572723001303> |
|  | Besdie pearson correlation and information gain, different evolutionary techniques like particle swarm optimization and co are utilized for feature selection. |  |  |  |  | <https://www.tandfonline.com/doi/full/10.1080/07853890.2023.2233541> |
|  |  |  |  |  |  | <https://www.jmir.org/2020/8/e20259/> |
| sudden shift in the data impacts xgb and nn model performances as well as feature importances |  |  |  |  |  | <https://link.springer.com/article/10.1186/s12911-023-02151-1> |
|  |  |  | 2 types of LSTM model are created. One with attention weights (for both temporal and non-temporal features) and one without attention weights. Since no attention module was presented in the second model, the outputs from the head and tail LSTM modules were merged with the non-temporal feature vectors and sent to a linear layer of 128 neurons with a sigmoid activation function as the classification module. |  |  | <https://www.frontiersin.org/articles/10.3389/fcimb.2022.838749/full> |
|  |  |  |  |  |  | <https://www.frontiersin.org/articles/10.3389/fpsyt.2021.752870/full> |
|  |  |  | The stacking ensemble model exploits a meta-learning model, which learns how to combine the prediction outputs of all basic classifers aimed at improving fnal classifcation efectiveness. |  |  | <https://www.sciencedirect.com/science/article/pii/S1568494620307973?casa_token=3zM8TVSf1zMAAAAA:ygxmUdJ38s0WArhW73_XdDzszvDmf7jXuMd24_Fk5WfUc0Gv41CyQyvyFhm-KOzj-UPYVORPc8c> |
|  | techniques to reduce the dimensionality or the number of variables, such as PCA (Principal Component Analysis), RFE (Recursive Feature Elimination) and SVD (Singular Value Decomposition), can eliminate redundant and irrelevant features. However, these methods may result in losses in the predictive power of ml methods.(3 |  |  |  | Gini index and information gain are both measures of impurity, but they are calculated differently. Gini index is calculated by summing the squared probabilities of each class, while information gain is calculated by comparing the entropy of the original set of data to the entropy of the two child sets. | <https://www.nature.com/articles/s41598-023-28579-z> |
|  | Boruta is an all-relevant feature selection wrapper algorithm, that can be used with any classification methods to produce the variable importance measure. Boruta iteratively compares the original features’ importance with the importance of permuted features (shadow attributes) to remove the irrelevant features, which are the features having significantly less importance as compared to the permuted features. features that are having significantly more importance than the shadow attributes will be confirmed as relevant features. |  |  |  |  | <https://www.medrxiv.org/content/10.1101/2023.01.17.22281858v1> |
|  |  |  | three ensemble models used the outputs of the three basic models (rf,xgb,lgb) as input variables for secondary prediction, the SA method gives each instance the same weight. The principle of OLS method is that the sum of the squares of the errors between the estimated value and the actual value is the smallest. The LAD ensemble computes forecast combination weights using the principle of minimum absolute errors. |  |  | <https://link.springer.com/article/10.1007/s11356-022-23132-3> |
|  | lasso regression is used as feature selection algorithm |  |  |  |  | <https://link.springer.com/article/10.1007/s00432-023-05417-3> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9767856?casa_token=ZIhcA4tSFaEAAAAA:GHtJCsyZ5Wmns1iikOhV10-ndjIRjvPr5Ul95SjguXBX1ED8aCoaW-YG87nfwGnm9qqM033Ijm9hMQ> |
|  | (1)For the statistics, analysis, features are divided into four groups—continuous normal, continuous non-normal, discrete binary, and discrete categorical and T-test, Mann–Whitney U-test, Chi-square test, and Fisher-exact test, are applied respectively.(2) RF Feature Importance implicitly accounts for the features that are both mutually correlated and correlated with the class. Information Gain, for example, would put all such features at the top of the list, despite the fact that they are mutually redundant.(3) the paper uses k-means clustering, to find subgroups of similar features. the input to the K-means algorithm was a transposed input matrix, where rows were features and columns were countries. Once the clusters of features were obtained, Principal Component Analysis (PCA) is used across all the features within each cluster as a dimensionality reduction technique to reducing the dimensionality to a single dimension. (\*) |  |  |  | To gain an insight into models, a problem-independent rule learning algorithm is created. The input to the algorithm is preprocessed by discretizing features into low and high values. The rule learning procedure evaluates all possible rules (, which is a computationally expensive procedure) with respect to the discretized features. Three metrics are calculated for each rule: (1) coverage, (2) accuracy on covered examples, and (3) rule score as coverage multiplied by accuracy. Rules with low coverage or low rule score on the training set are discarded. The remaining rules are sorted. | <https://www.mdpi.com/1660-4601/18/13/6750> |
|  |  | hyperparameters were set using Bayesian optimization |  |  |  | <https://www.medrxiv.org/content/10.1101/2020.05.07.20094573v2> |
|  | (1) an unsupervised clustering approach combining the Gower distance with the Partitioning Around Medoids (PAM) algorithm k-Medoids is applied to identify three clusters differing significantly in subjects’ well-being, psychological distress, and current financial and occupational concerns.(2)Features with the strongest impact on the clustering were examined using a wrapping method and the feature importance implemented in the random forest. |  |  |  |  | <https://link.springer.com/chapter/10.1007/978-981-16-2377-6_81> |
|  |  |  |  |  |  | <https://www.researchsquare.com/article/rs-2556795/v1> |
| (1) variance inflation factor and Pearson’s correlation analysis are used for feature selection(2) |  | optimum number of neurons in the hidden layer is determined using WIC index. Based on this method, the number of neurons in the hidden layer was systematically increased from one to the number of variables, and then the WIC index value of each model was calculated. The lower model’s WIC index indicates a more efficient model.(3) |  |  | 10 feature importance methods are used: cw, mcw, (which are calculated based on the sum of product of final weights of connections from input neurons to the output neurons), most square (based on the sum of the squared between initial weight and final weight), Garson’s algorithm(),pd as well as stepwise and perturb methods. | <https://www.sciencedirect.com/science/article/pii/S1877584521000691?casa_token=iqKe9fsvkcUAAAAA:4wQGq8z_3_sApG2SXp09Cik0gws826p96uNKJJlESYcvoGNDOcnbfyFVU30E35uBTap6kYZlOsM> |
|  |  |  |  |  |  | <https://books.google.de/books?hl=en&lr=&id=1PGiEAAAQBAJ&oi=fnd&pg=PA57&dq=covid+intext:%22explainable+machine+learning%22&ots=OTPYyTnfQZ&sig=D3UTEo9qwdg1ZN57SbkAj29-tvI#v=onepage&q=covid%20intext%3A%22explainable%20machine%20learning%22&f=false> |
|  | (1) the WHO dataset was enhanced with thirty different target classes, which were obtained through a machine learning-based clustering method. k-means and PCA method is used to identify patterns in data and reduce the number of features while maintaining important information. Latent Dirichlet Allocation (LDA) for topic modeling is employed to discover topics of the scholarly documents in each cluster. LDA represents each scholarly document as a distribution of topics and each topic as a distribution of keywords. the keywords identified as the most probable terms to be associated with each topic were used to define the topic clusters. Specifically, the first three keywords were chosen as the cluster name based on their high probability of being associated with that topic. |  |  |  |  | <https://www.researchsquare.com/article/rs-2801916/v1> |
|  |  |  | The paper introduces 6 HER specific predictive models, e.g. Dr.Agent and Retain, which, often apply the attention mechansim and their performance surpasses that of basic deep learning models, which in turn outperform traditional machine learning models. Retain Reverse Time Attention model is an interpretable deep-based model for analyzing EHR data. It utilizes a two-level neural attention module to attend important clinical visits and features. Dr.Agent augments RNN with 2 policy gradient agents. It learns a dynamic skip connection to focus on the relevant information over time. |  |  | <https://arxiv.org/abs/2209.07805> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/abs/pii/S0957417423014409> |
|  |  |  |  |  |  | <https://www.researchsquare.com/article/rs-90375/v1> |
| To create a distribution and confidence intervals of the model performance, as performance may change depending on the choice of split, multiple non-random splits by time were created. s it allows for non-random variation between the train and test sets, since all records of the test data of each split come from a time window which has not been seen during training of the respective split. A sliding window was applied on the dataset to create 100 different splits, where the window of 20% width corresponded to the test set of the split. Thus, for the first split, the test set covered the first 20% of the data records (earliest cases), while the test set of the 100th split corresponded to the last 20% (most recent cases). The non-binary features of the train set and the test set were imputed based on multivariate feature imputation using Bayesian Ridge estimation. | For each split, the features were ordered by decreasing mean absolute SHAP value and only the features representing when combined 95% of the sum of the mean absolute SHAP values were retained, thus removing the least important features which combined contribute to less than 5%. The model is trained for each split. The final XGBoost model was retrained on the entire set of records with the reduced set of features. |  |  | To tune the XGBoost hyper parameters, a five-fold cross-validation grid search on the training data maximizing the ROC AUC was used. |  | <https://www.nature.com/articles/s41746-021-00482-9> |
|  |  |  |  |  | Identifying these features constitutes an important step in the debugging process, which should ensure that the model has learned to look at the pixels (features) that convey the appropriate information to classify the image. It might occur that, in the training process, a model trained to associate certain patterns with the labels can actually mislead the classification process. For instance, one can train a model to distinguish images with cats labeled as “CATS” from images with dogs labeled as “DOGS”. If the images with cats were all obtained outside on sunny days and the images with dogs were all taken indoors, the model could wrongly learn that blue skies should be associated with the label “CATS” and that the lack of blue skies should be linked to the label “DOGS”. If the testing set was obtained from the same set of images—i.e., the testing images of cats also have blue skies and the images of dogs are indoors—the model will likely correctly classify the cats as “CATS”, but only because of the blue skies rather than because of the cats shown in the images. Likewise, it will classify images with dogs as “DOGS” due to the features pertaining to the indoor environment. However, in this case, the same model would probably classify an image of a dog walking in a park on a sunny day as “CATS”. Therefore, it is important to confirm that the features (pixels) that a given model actually uses to classify images are appropriate. | <https://www.mdpi.com/1999-4893/14/6/183> |
| The SMOTE algorithm draws a random sample from the minority class, identifies the k nearest neighbors, takes one of those neighbors and identifies the vector between the current data point and the selected neighbor, multiplies the vector by an appropriate random number, and finally adds this to the current data point |  |  |  |  |  | <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0276767> |
| The paper uses FDA to obtain data with a higher signal-to-noise ratio and more accurately constructs the intrinsic public emotion to better investigate the dynamic evolution of emotion. Functional data analysis fda follows the “breaking up the whole into pieces” principle of big data analysis, can handle discrete and high-frequency sequences of data as continuous smooth functions treating the whole functions as a single entity with an internal unified structure. |  |  |  |  |  | <https://www.mdpi.com/1660-4601/19/20/13248> |
|  |  |  |  |  |  | <https://www.mdpi.com/2504-2289/6/4/106> |
|  |  |  |  |  |  | <https://www.emerald.com/insight/content/doi/10.1108/IDD-01-2021-0003/full/html?casa_token=bOUCJu-5HkIAAAAA:8I3WoU89vU4Ui5QAOVvzsd-0_JNJJtVJAXKyQoKI95RX1PuBGqx5zDG9vjpc-Cavz81nYdf9sVlMiwK_WBELOXbg-HeoLiw5X3ApKbl7LltWJkJZPd2I> |
|  |  |  |  |  | the effectiveness of deep models for COVID-19 diagnosis from CTs is improved by attention-based strategies allowing a preliminary extraction of most informative sub-volumes, used for patient classification. By focusing on the most infected slices, the model complexity is reduced, so that abrupt and lossy image reduction can be avoided. | <https://www.mdpi.com/1660-4601/18/6/2842> |
| predictive mean matching to impute numeric features, logistic regression to impute binary variables and Bayesian polytomous regression to impute factor features are used. | LASSO regression is used to select and sort the statistically significant clinical features. |  |  |  |  | <https://www.nature.com/articles/s41467-020-17280-8> |
|  |  |  |  |  | The filtered outputs are pooled, log-compressed and used in a self-attention based relevance weighting mechanism. The relevance weighting emphasizes the key regions of the time-frequency decom[1]position that are important for the downstream task. | <https://arxiv.org/abs/2206.13365> |
|  |  |  |  |  |  | <https://www.mdpi.com/2075-4426/12/2/309> |
|  | The Point-Biserial Correlation is a special case of the Pearson Correlation and is used when we want to measure the relationship between a continuous variable and a dichotomous variable, or one that has two values |  |  |  |  | <https://link.springer.com/article/10.1007/s00500-022-06943-x> |
|  |  |  |  |  | Dual stage attention DA-RNN model involves two attention layers: the first is an input attention layer that determines which feature should be given more attention than the others; the second is a temporal attention layer that determines the weight of importance for each historical temporal step. | <https://link.springer.com/article/10.1007/s10489-021-02616-8> |
|  | Relevant features are selected using the Tversky similarity-indexed stochastic distributive embedding technique. |  |  |  |  | <https://www.proquest.com/openview/129cfa97dcf3170a6a25bae5e57e98cf/1?pq-origsite=gscholar&cbl=2037694> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9724545?casa_token=HDJ27DEqzYQAAAAA:nF4y6uBXBgoqBYvNo9NJgCOId8D8rL6sRxf3LMEdklD6mGFzPRrQb4yvVzJL5ANnA8OhrZuzdWi73g> |
|  |  |  |  |  | (1)Feature attention modules are used to encourage CNN to learn and focus on task-relevant information instead of learning non-useful information | <https://www.mdpi.com/2075-4418/11/2/158> |
|  |  |  |  |  |  | <https://link.springer.com/article/10.1007/s11604-023-01466-3> |
| (1)As base imputation models, Predictive Mean Matching (PMM)-mice19 and Random Forest classifiers (RF)-mice are used. PMM-mice essentially uses the values of a neighbor, where the neighbor is randomly selected among the k nearest points in bootstrap samples |  |  |  |  | .(2) the random forest model was then pruned and simplified to create a simple associative tree. | <https://www.tandfonline.com/doi/full/10.2147/RMI.S292314> |
|  |  |  | (1)a pre-trained convolutional neural network (convnet) model is employed as a feature extractor. After extracting features from the CXR and CT scans individually, all of the data from each modality are aggregated and feed the combined features to a new DL classifier(2) |  |  | <https://ieeexplore.ieee.org/abstract/document/9671302?casa_token=ZkwUxHiTyRsAAAAA:qt4eep2zrgGiHfhwWECfp8zQs_e4semt_xwQ4tHySlSrdAtz7fIijfowTgo1ie3IMJz81VZPICdR1w> |
|  | research into detection of disease from portable CXRs is challenging due to the low resolution. Hazy pulmonary opacities in CXR can sometimes be blurred which makes detection more difficult. In order to use DL algorithms to handle such hazy regions and irregularities of portable CXR images, using a combination of CXR and fuzzy features, can result in a higher classification performance. |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9882738?casa_token=H7HyR-m6-c8AAAAA:uTmaNklF3GY4Byg6SQzRv_024qWoNmYWAz4I3_nApaB5BDr3NpJbTCJ49nchnx_Zvf8SdDAG1zMp0w> |
|  |  |  |  |  |  | <https://academic.oup.com/eurpub/article/33/5/930/7226726?login=false> |
|  | feature selection is about selecting the subset of the original feature set, whereas feature extraction creates new features |  |  |  |  | <https://link.springer.com/chapter/10.1007/978-981-19-0098-3_31> |
|  |  |  | To understand Geospatial Evolution of COVID-19, Methods like seir, var, arima, lstm consider the input data as independent sequences of vectors and do not explore the spatial context available in Spatio-temporal data. In Convolutional LSTM (ConvLSTM) approach, the input data structures are 3D tensors, with the first dimension corresponding to the number of relevant measurements and the last two dimensions representing the spatial dimensions (i.e., width and height). |  |  | <https://dl.acm.org/doi/full/10.1145/3550272> |
|  | (1)the feature map is the output of one filter applied to the previous layer(2) SOM groups the data to a number of small and compact clusters with respect to BMU best matching unit taking the neurons as the cluster centers |  |  |  |  | <https://bcas.edpsciences.org/articles/bcas/abs/2022/01/bcas2022003/bcas2022003.html> |
|  |  |  |  |  |  | <https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000100> |
|  | Dynamic Time Warping algorithm permits identification and classification of time series with similar longitudinal patterns |  |  |  |  | <https://www.researchsquare.com/article/rs-3446631/v1> |
|  |  |  |  |  |  | <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9840564/> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/abs/pii/S1059056023003039> |
|  |  |  |  |  |  | <https://europepmc.org/article/med/37474793> |
|  |  |  |  |  |  | <https://journals.eanso.org/index.php/ajthm/article/view/1522> |
|  |  |  |  |  |  | <https://modelling.semnan.ac.ir/article_7795.html?lang=en> |
|  |  |  |  |  |  | <https://iariw.org/wp-content/uploads/2022/08/Romanko-OMahony-IARIW-2022.pdf> |
|  |  |  |  |  |  | <https://www.researchgate.net/profile/Juracy-Bertoldo-Santos-Junior/publication/353231152_Inteligencia_artificial_e_COVID-19/links/61f2b8bfdafcdb25fd55cc5f/Inteligencia-artificial-e-COVID-19.pdf> |
|  |  |  |  |  |  | <https://repositorium.hs-ruhrwest.de/frontdoor/index/index/docId/782> |
|  |  |  |  |  |  | <https://www.maxwell.vrac.puc-rio.br/60658/60658.PDF> |
|  |  |  |  |  |  | <https://sol.sbc.org.br/index.php/sbcas/article/view/21627> |
|  |  |  |  |  | (1)posthoc visualization methods fail to explain the actual reasoning process behind the classifications made by the models. ()2)a prototypical part represents a part, a patch or a section of an image. The similarity score between a learned prototypical part and an image is considered as the sum of squares of the differences between the corresponding pixel values of the prototype and a patch of the image. To predict the class of a test image, Gen-ProtoPNet calculates the similarity scores between learned prototypical parts of images from each class and parts of the test image. In the cnn, the convolutional layers are followed by a prototype layer, which is in charge of computation of similarity scores. | <https://ieeexplore.ieee.org/abstract/document/9448270> |
|  |  |  |  |  |  | <https://www.nature.com/articles/s41746-021-00496-3> |
|  |  |  |  |  |  | <https://www.mdpi.com/2075-4418/11/9/1732> |
|  |  |  |  |  | TabNet exploits the benefits of sequential attention to choose features at each decision step and utilizes a feature value output called Masks, which shows whether a feature is selected at a given decision step in the model and can be used to calculate the feature importance. The masks can be represented by heat maps for each input feature represented by each row, and a sample from the data set represented by columns. | <https://peerj.com/articles/cs-889/> |
|  |  |  |  |  |  | <https://www.spandidos-publications.com/10.3892/etm.2020.8797> |
|  |  |  |  |  | the approach adds multiple attention layers in a hierarchical manner to obtain first the higher levels attention scores of images. This Contributes to increasing sensitivity of the model by identifying the specific blocks within each individual’s chest image that may include subtle signs of infection as well as providing interpretability to the results. | <https://dl.acm.org/doi/abs/10.1145/3466690?casa_token=bQYxuppO-_cAAAAA:5ExUWvmPdvpa9RSU7j8sYTnf86x2PXiwHK9lO8CrlISxEP3-9mlSHLbJqKlq4kbwenOsamrukM7c1Q> |
|  |  |  |  |  |  | <https://www.medrxiv.org/content/10.1101/2020.05.16.20103408v2> |
|  |  | Standard deep learning models employ the categorical cross-entropy loss to perform image classification tasks. This work shows that incorporating the orthonormality loss produces improved semantic localization via GradCAM visualizations. |  |  |  | <https://arxiv.org/abs/2102.08360> |
|  |  |  |  |  | This however gives rise to the credibility of attained explanations as interpretability analysis derives from a different modelling process with priors that are not part of the training from the original networks.(2) Inspired from case based reasoning and autoencoder architecture, this study comprises a network, which builds a profile layer comprising a list of profiles (or prototypes) whereby each profile resembles observations in one of classes in visual appearance. Hence this set of profiles learns toward being a representative of the whole training set. when a test image is loaded to the trained model, the model calculates the overall distance between this test image and each of the profile images and delivers the final classification result. | <https://ieeexplore.ieee.org/abstract/document/9680088?casa_token=DN8VB7Q9cpoAAAAA:0_RQGpQkQF8zaQlaIqSz40nQU1haVzpJfR1szvSHF6NzI3Ymk9CkOIPJUhyUy5jj2Zb2HReRdw7bkA> |
|  |  |  |  |  | 2D Convolutional Neural Networks (CNN) have been primarily used for computer vision applications, where multiple filters are trained to detect different input image features. 1D CNNs have been shown to work on time series problems like longitudinal EHR data. 1D Convolution works over the temporal dimension with different filter sizes, where the different filters learn different temporal patterns. This process produces feature vectors which are then passed through a nonlinear layer like a Rectified Linear Unit (ReLU) or Tanh. Gradient-weighted Class Activation Mapping (GradCAM) has been used to examine 1D CNN based models that analyze protein sequences and find regions in the input sequences that help the model make the correct prediction. GradCAM is generally used in computer vision to generate localization maps for a given concept (class) in the input image. These maps are made by finding the gradient of the predicted class in the activation map of the final layer, pooling them channel wise, and the resultant activation channels are weighted with the corresponding gradients which can then be inspected to find which parts of the input helped in the classification. | <https://ieeexplore.ieee.org/abstract/document/9994851?casa_token=sPuQ7pr0YEYAAAAA:Q2smeXVFJ7aHAhW-JLVR5PVjcFqTNOdaiw3Hn1u0RxQSmxJC1P2ePub-pYKoyQ3MMRFQLQ7tdckgTw> |
|  |  |  |  |  | (1)In the gamut of machine learning algorithms, regressions remain one of the simplest and most explainable models with clear mathematical formulation.(2) the paper addresses the symbolic meta modelling, which integrates various simple parameterized functions to get a closed-form and interpretable expression for the meta model. | <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12088/120880A/Global-and-local-interpretation-of-black-box-machine-learning-models/10.1117/12.2604743.short?SSO=1> |
|  |  |  | (1)a multi-stage (4 stages) lstm deep learning framework is proposed, which, at each stage, forecasts a chosen target variable for the seven days ahead (e.g., one-week ahead forecast). The multi-stage model builds off the initial first stage prediction to forecast an additional week out and continues to implement this iterative approach one stage at a time, to predict further into the future.(2) | the model performance is compared with CDC COVID-19 Forecast Hub model, which compile multiple models of various approaches within a single prediction framework and has proven to be the best performing approach for short term COVID-19 forecasting efforts. | IG integrated gradients method assigns importance to input features as attributions. It achieves this by integrating the gradients of the output with respect to the input along an arbitrary path from the baseline to the input data point. The baseline could be e.g in the case of NLP a zero embedded word vector or a black image in computer vision.(3) | <https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(23)00047-6/fulltext> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0010482522004498?casa_token=Ai61Zxzn0bwAAAAA:6wJNrNGbhqrAYAisss9_uLp20ssn5FfOrlxxLSUfmPBAuMJ0--LTEDYMHigwkZxjFYKSMQjkHoY> |
|  |  |  |  |  |  | <https://www.cell.com/patterns/pdf/S2666-3899(22)00002-2.pdf> |
| A novel self-adaptive auxiliary loss is proposed to help the training with imbalanced data by introducing a self-adaptive factor, which reflects the feature distribution and emphasizes minority class. |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0925231221009115?casa_token=LsikVMZ4hLAAAAAA:vDjBqxE-LhlKLGe4i50kMyUKNdzOWpwdO8DP5-cACbnz-cqTPrfJGTIxEOpfyBLc3e3s_bTV8Wg> |
|  |  |  |  |  | For sentiment analysis, the use of attention e.g. to focus on an essential part of the sentence and convolution neural networks has emerged as a popular strategy. | <https://www.sciencedirect.com/science/article/pii/S0378720621001610?casa_token=eR6zxPle8DwAAAAA:0-U8EH1hcZmy144Z-99IWaEALwp9x_p1lCnRc8CGmt_HhRpA8VkEDLNrwOoeYWy3k7W47GpJsSM> |
|  |  |  |  |  | feature ablation investigates the performance of an AI system by removing certain components to understand the contribution of the component to the overall system (\*)(1)behrt is a transformer model for handling HER records(2)The behrt’s self attention mechanism gives it the ability to find the relationships among events which goes beyond temporal/sequence adjacency. | <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9285184/> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9910274> |
|  |  | splitted rolling windows from residual time series of various regions are taken and are transformed into a common feature space using normalized convolution as embedding functions. attention mechanism are employed to measure the relation of one region to other regions by directly comparing the incidence curves after trend filtering. In such pairwise comparisons, differences in both time-dependent and time-independent features are taken into account so that the curves in corresponding windows can be better aligned. |  |  | attention mechanism is used, to learning inter-series similarity to generate time series forecasting. | <https://epubs.siam.org/doi/abs/10.1137/1.9781611976700.56> |
|  |  |  |  |  | The deep network corresponded to each of two critical domain’s contribution to the final prediction are captured by a corresponding coefficient of a linear regression. | <https://ojs.aaai.org/index.php/ICWSM/article/view/19353> |
|  |  |  |  |  |  | <https://publish.kne-publishing.com/index.php/fbt/article/view/5321> |
|  |  |  |  |  | a qualitative comparison between different association rule mining algorithms is presented in this paper. | <https://www.sciencedirect.com/science/article/pii/S0010482521000433?casa_token=GuzO7A9P4sYAAAAA:s-orPejv7NZ1i3WwtIcCtMSbdqf-QL6mXVH7831LHxqKssr8exyf6qI_amqo0wb9OiylVa4Eq7A> |
|  |  |  |  |  | Multi-resolution approaches, which observe and evaluate a dataset at several resolutions, are popular in the machine learning domain. They capture the local geometry of neighbourhoods, which are characterized by a collection of distances between points or groups of closest neighbours. This is analogous to looking at a portion of a slide at various microscopic resolutions; whereby, very small features can be detected at high resolution from a restricted region of the sample. As the majority portion of the slide is examined at a lower resolution, it allows one to examine the larger (global) features as well. | <https://www.sciencedirect.com/science/article/abs/pii/S0957417423016019> |
|  |  |  |  |  |  | <https://www.igi-global.com/chapter/an-effective-video-surveillance-system-by-using-cnn-for-covid-19/287230> |
|  |  |  | the paper presents a hybrid lstm-nn model. Time series data is represented as a 2D vector of sequences, with 14 days of each of the features. This input is processed using two Keras LSTM layers with 120 neurons each, separated by a dropout layer. Static data is passed through a single simple neural network layer (a Keras Dense layer) containing three neurons. This dropout layer was added to help prevent overfitting of the network. A final layer produces the output of the neural network architecture. |  |  | <https://ieeexplore.ieee.org/abstract/document/9581192?casa_token=6YMR1-E-e4AAAAAA:1jMrpnfiYOx10QsUKia39ukJMsEBZ-vwS_Kf2iytZ2qhGbH5i07bUEuldAFZztdXFlysG-HdqTgyJg> |
|  |  |  | This paper builds time series (via lstm) of graphs that reflect the underlying connections of atmospheric conditions across local regions. That is, attaching weighted edges as a function of temporal-varying temperature or relative humidity provides a natural way to track weather changes between regions and link these with the trend in COVID-19 dynamics. |  |  | <https://link.springer.com/chapter/10.1007/978-3-030-75762-5_17> |
|  |  |  |  |  |  | <https://www.mdpi.com/2079-7737/12/6/887> |
|  |  |  |  |  |  | <https://link.springer.com/chapter/10.1007/978-981-99-1410-4_29> |
|  |  |  |  |  | To assess the effect of each individual machine or deep learning model, an ablation approach is used to estimate the accuracy of ensemble with and without each base model. | <https://ojs.aaai.org/index.php/AAAI/article/view/26855> |
|  | Third, the aane Accelerated Attributed Network Embedding (to extract the low-dimensional representation of the network) algorithm was adopted to capture the typological features as well as node attributes. Next, clustering analysis (K-means and Gaussian mixture) was performed on the attributed network embeddings every week. During this period, clustering results were computed, and counties with the same clustering tendencies were merged into archetypes. Finally, Kruskal Wallis and Dunn’s tests were conducted to explore the feature importance i.e. archetype differences and expose features underlying distinct transmission risk patterns across archetypes. |  | The attribute network embedding model created with two main components: (1) county-level features (node features); and (2) cross-county movement network (spatial network topology). The attribute matrix was obtained in the first step and was calculated into the attribute similarity matrix. Second, a spatial network structure was constructed based on cross county movements and was then expressed as an adjacency matrix. |  |  | <https://arxiv.org/abs/2209.09448> |
|  |  |  |  |  |  | <https://www.mdpi.com/2076-3417/11/15/7080> |
|  | Knowledge graph embedding is a technique to encode the entities and relations in a Knowledge graph as dense and low-dimensional vector representations. |  |  |  |  | <https://link.springer.com/article/10.1007/s10115-023-01923-5> |
|  |  |  | the model builds an agent based model network structure based on the colocation of phone numbers. On top of the network a seir model is applied |  |  | <https://www.sciencedirect.com/science/article/pii/S026427512200244X?casa_token=iFYfvBXdJfUAAAAA:PADxApsJq0TlzqKlFUHIuAMu3zj8f4UMST90UOpTRPztf9-6Dlv826Ohj3hMGrbWqH6p8gRZo_U> |
|  |  |  |  |  |  | <https://appliednetsci.springeropen.com/articles/10.1007/s41109-021-00361-y> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0269749122006558?casa_token=nS4RU0vW-fAAAAAA:fSn6fkGq3u_UsO3L1BZ327xRLnvoqvs4M05tWC94w4RJVwjYJWcmecdV0L5OZMha9RdO5aQMJwY> |
| multiple GAN architectures and their ability to generate realistic synthetic samples of COVID-19 chest X-rays to deal with limited numbers of training samples are evaluated. |  |  |  |  |  | <https://link.springer.com/article/10.1007/s00521-020-05636-6> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/10215391?casa_token=vrxTAV3UM-0AAAAA:z4jCu7dFDP_S2fLNqPKusTuySddzUqRXJPSldgU7Abrrm7zIqvJp3Yg54FDmToU81XEU_sZfbTOTJA> |
|  |  |  |  |  | A figure depicts Motivation for model developers and clinicians to implement XAI in AI-enabled clinical decision support systems. | <https://www.nature.com/articles/s41598-023-46493-2> |
|  |  |  |  |  | Paper proposes and assesses three techniques’ explanation techniques: Composite Layer-wise Relevance Propagation, Single Taylor Decomposition, and Deep Taylor Decomposition, which are based on the extension of the idea of taylor expansion to the set of neurons in neural nwtorks. The idea of Deep Taylor Expansion is to understand that a complex and non-linear function can be decomposed into a set of simpler sub functions. | <https://link.springer.com/article/10.1007/s00521-022-08021-7> |
|  |  |  |  |  |  | <https://link.springer.com/article/10.1007/s11063-021-10712-6> |
|  |  |  | FKNN fuzzy classifier improves upon the conventional KNN clas\_x0002\_sifier by adding the concept of fuzzy logic to the KNN |  |  | <https://ieeexplore.ieee.org/abstract/document/9420727/> |
|  |  |  |  |  | prisma | <https://link.springer.com/article/10.1007/s44196-023-00236-3> |
|  |  |  |  |  | this paper proposes ‘Explanation Score,’ a weighted overlap score between the lesion region and positions of decisive features. | <https://ieeexplore.ieee.org/abstract/document/9684653?casa_token=b75jj8cQLR8AAAAA:lrTaPa4IZNSYsVGccP2TAKuRdT2SliwSuSdFbElckorJEcoO7kVpJGqv_-i4sJkZcqxy3pbFJHDedQ> |
| (1)the study utilizes a time series augmentation technique to create new series with the same temporal dependencies that exist in the original series. The augmented time series is used to create a new validation set to fit deep learning model more properly. |  |  |  |  |  | <https://link.springer.com/article/10.1007/s00521-021-06548-9> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S1532046421001209> |
|  |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S1568494620308449?casa_token=ZiKJFj1lFQkAAAAA:A8sjj3quw9Zszw4SBSzIZ0yJl_CSMSJmzp4T9zQTWe_GCWg_lpfJI40aCkHjFt2ejBqFv4kI8sQ> |
|  |  |  | The classification model is a hybrid model that consists of two classifiers; fuzzy inference engine and Deep Neural Network (DNN). Fuzzy Logic (FL) describes systems in terms of a combination of numeric and linguistic (symbolic). This has advantages over pure mathematical (numerical) approaches or pure symbolic approaches because very often system knowledge is available in such a combination. |  |  | <https://www.sciencedirect.com/science/article/pii/S1568494620308449?casa_token=ZiKJFj1lFQkAAAAA:A8sjj3quw9Zszw4SBSzIZ0yJl_CSMSJmzp4T9zQTWe_GCWg_lpfJI40aCkHjFt2ejBqFv4kI8sQ> |
|  |  |  | BiLSTM-A-CNN, is intended to amalgamate the ability of LSTMs to model time dependencies, the capability of the attention mechanism to highlight relevant features, and the noted ability of CNNs to extract features from complex data. The use of the BiLSTM-A-CNN model is expected to improve the forecasting accuracy of future |  |  | <https://iopscience.iop.org/book/edit/978-0-7503-3795-3/chapter/bk978-0-7503-3795-3ch5> |
| generative adversarial network was created to increase the COVID-19 dataset. |  |  |  |  |  | <https://www.sciencedirect.com/science/article/pii/S0957417421008241> |
|  |  |  |  |  | J48 is a machine learning decision tree classification algorithm based on Iterative Dichotomiser | <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8719570/> |
|  |  |  |  |  |  | <https://www.researchgate.net/profile/Saif-Al-Jumaili/publication/370956355_Evaluation_of_Deep_Transfer_Learning_Methodologies_on_the_COVID-19_Radiographic_Chest_Images/links/646f15ebcde97a392931d45e/Evaluation-of-Deep-Transfer-Learning-Methodologies-on-the-COVID-19-Radiographic-Chest-Images.pdf> |
|  |  |  |  |  |  | <http://journals.ajsrp.com/index.php/jesit/article/view/5284> |
|  |  |  |  |  |  | <https://www.mdpi.com/2075-4418/13/8/1466> |
|  |  |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9336835> |
|  | the paper adopted two-step strategy, where filter measure is applied to rank the features according to their relevance in the first step, and Genetic Algorithm is applied with Decision Tree classifier to find the best feature subset in the second step. |  |  |  |  | <https://ieeexplore.ieee.org/abstract/document/9628327> |
|  |  |  | the study utilizes two knowledge graph based ML models, logistic regression (LR) and random forest (RF) models. The COVID-19 knowledge graph (KG) constructed based on literature from heterogeneous data is imported to understand the COVID-19 different relations. |  |  | <https://peerj.com/articles/cs-1333/> |
|  |  |  |  |  |  | <https://www.mdpi.com/2075-4418/12/12/3171> |
|  |  |  | this paper uses Fuzzy Decision Tree (FDT), and Fuzzy Random Forest (FRF) |  |  | <https://www.mdpi.com/2227-7390/9/24/3282> |
|  |  |  | propose a hybrid framework that combines a Convolutional Neural Network (CNN) with a fuzzy system to identify infection severity |  |  | <https://ieeexplore.ieee.org/abstract/document/10278298> |