|  |
| --- |
| Table 1. Summary of enzyme engineering approaches and new trends |
| Type of improvements | Brief about the advancements  |
| Workflow/pipeline: for directed evolution (DE). | A high-throughput investigation to optimize the machine learning assisted directed evolution (MLDE) workflow by testing a number of design considerations of the MLDE pipeline available at <https://github.com/fhalab/MLDE> ([Wittmann et al. 2020](#_ENREF_26)). |
| A predictive DE model.  | An innovative sequence-activity relationship (innov’SAR) methodology based on digital signal processing combining wet-lab experimentation and computational protein design. This model is based on An epoxide hydrolase from *Aspergillus niger*([Cadet et al. 2018](#_ENREF_5)). |
| A predictive model for catalytic turnover number (*k*cat). | The model has identified a diverse set of features such as structure, biochemistry, network those are applicable for *in vivo* and *in vitro* enzyme turnover rates to correlates with the predictive catalytic turnover rates ([Heckmann et al. 2018](#_ENREF_13)). |
| A predictive model for optimal growth temperatue, catalytic temperature optima (*Topt*). | This model is used to generate the optimal enzyme *Topt* which helps to redesign enzymes that are potentially functional at extreme temperature ([Li et al. 2019](#_ENREF_14)). |
| A predictive model of concentration for metabolic flux optimization. | The model is based on the artificial neural network and it could be useful for optimization of *in silico* enzyme concentration prediction for the cell free enzyme assay ([Ajjolli Nagaraja et al. 2020](#_ENREF_1)). |
| Machine learning sequence function models and to use those models.  | This high-quality review work provides steps for machine learning sequence function based models for accurate protein engineering through directed evolution ([Yang et al. 2019](#_ENREF_28)). |
| Machine learning- based improvement of proteinase K. | In this model, about 20X improvement of the catalytic efficiency was achieved by using two cycles of machine learning algorithms by testing only 95 variants of redesigned proteinase K ([Liao et al. 2007](#_ENREF_17)). |
| Supervised machine learning based ligand affinity predation models. | This work provides detailed information on supervised machine learning-based ligand affinity prediction methods that could be useful for enzyme engineering ([S Heck et al. 2017](#_ENREF_24)). |
| An ensemble learning model for accurate prediction of the optimum catalytic temperature (*Topt*) of the enzymes. | It is an improved ensemble learning model for eliminating error in prediction temperature range of the enzyme ([Gado et al. 2020](#_ENREF_12)). |
| Machine learning based prediction model for enzyme activity and substrate specificity of thiol superfamily enzyme. | This model is based on enzymes in the thiolase superfamily and measured the activity of 73 diverse bacterial thiolases. This model is available at <https://github.com/serina-robinson/thiolase-machine-learning/>([Robinson et al. 2020](#_ENREF_22)). |
| A high-quality and high-throughput deep learning model for accurate enzyme commission (EC) number prediction model. | This is a high-precise deep learning model that is based on three convolutional neural network and homology analysis for the prediction of the EC number ([Ryu et al. 2019](#_ENREF_23)). |
| A multi-level machine learning model enzyme-substrate prediction. | This model is based on experimental activity data, structural modeling, ligand docking, physiochemical properties of protein ligand based on bacterial nitrilase ([Mou et al. 2020](#_ENREF_19)). |
| A multi-level hierarchical deep learning model for multi-functional enzyme prediction. | This deep learning model is based on a novel loss of function associated with the relationship between different levels and self-adapted level assigning threshold ([Zou et al. 2019](#_ENREF_30)). |
| Proposed machine learning model for class selective optimization of enzyme. | This work emphasised the application of machine learning for practical improvement of biotechnology, metabolic engineering, and synthetic biology ([Ng 2020](#_ENREF_20)). |
| Directed evolution model of the enzyme- based on a statistical exploration of sequence-function space. | This report provides the usefulness of machine learning assisted directed evolution of enzymes rather than enzyme optimization by random mutagenesis, DNA shuffling, *etc*. ([Fox and Huisman 2008](#_ENREF_11)). |
| Automatic single, multi-level enzymatic function prediction model. | The model combines both structural, and amino acid sequence information sequence. The approach also includes features level and decision level investigation for accurate prediction of enzyme commission number. This machine learning model is available at <https://figshare.com/s/a63e0bafa9b71fc7cbd7> ([Amidi et al. 2017](#_ENREF_2)). |
| A machine learning model for identification of the reactivity promoting region (RPR) of the enzyme. | The model is based on few important descriptors such as substrate conformation, metal coordinate geometry, and substrate bond polarization to promoting the substrate reactivity with < 85% accuracy ([Bonk et al. 2019](#_ENREF_4)). |
| A Random Forest -based machine learning model for enzyme reaction prediction.  | This prediction model utilizes two-fold accuracy optimizations by predicting Enzyme Commission (EC) number from sequence data, and secondly enzyme-substrate models ([Watanabe et al. 2020](#_ENREF_25)). |
| A supported vector machine (SMV) model for substrate specificity prediction. | This SMV model uses a large set of data, and approximately 80% accurate prediction with approximately 30% less compound in the datasets. ([Pertusi et al. 2017](#_ENREF_21)). |
| A quantitatively validated machine learning model for enzymatic pathway prediction. | A large data set was used for 123 pathway feature collection, and used as an input to the extensive collection of machine learning methods, such as decision tree, logistic regression, *etc.* ([Dale et al. 2010](#_ENREF_8)). |
| A multi-level machine learning model for prediction of the enzymatic mechanism.  | The model utilise the large set of databases from InterPro, Catalytic site Atlas, MACiE, EzCatDb, SFLD using off-the-shelf K-Nearest Neighbours multi-label algorithm available online at <http://sourceforge.net/projects/ml2db/> ([De Ferrari and Mitchell 2014](#_ENREF_9)). |
| A high-performance machine learning- based tool for metabolic pathway prediction of plant enzymes.  | This model is based on sequence similarities of the enzymes based on the reference sequence, and it is available for local installation using a Graphical user interface ([de Oliveira Almeida and Valente 2020](#_ENREF_10)). |
| A hyper network model for enzymatic weight update. | The molecular algorithm is based on training data and by targeting internal loop structures in DNA and ensemble learning ([Baek et al. 2019](#_ENREF_3)). |
| Multiple machine learning algorithms for prediction enzymatic reactions.  | Three reaction fingerprints, and seven machine learning models were used for the prediction of the enzymatic reaction catalysed by oxydoreductase, and hydrolase ([Cai et al. 2018](#_ENREF_6)). |
| Supervised machine learning based enzyme class prediction.  | The enzyme class prediction is based on amino acid sequence derived features, such as composition, dipeptide composition, distribution, *etc.* Support vector machine recursive feature elimination and Random Forest are used for prediction ([Yadav and Tiwari 2015](#_ENREF_27)). |
| An online server for enzyme selective pathway design. | “Selenzyme” is an assembled tool with the extended application of multiple tools such as machine learning, antiSMASH *etc.* available at <http://selenzyme.synbiochem.co.uk/> ([Carbonell et al. 2018](#_ENREF_7)). |
| A semisupervised Gaussian model for enzyme search, *Km*prediction. | The automated enzyme search is based on chemical transformation fundamentals by a semisupervised gaussian model to provide a probability estimates. Moreover, these estimates were validated experimentally in *E.coli* ([Mellor et al. 2016](#_ENREF_18)). |
| Machine learning models for metabolic engineering.  | This work illustrates how machine learning models can use to overcome the rate-limiting step and for optimization of convoluted metabolic networks ([Zhou et al. 2020](#_ENREF_29)). |
| A deep learning model for accurate enzyme function prediction.  | DEEPre, is a deep learning model based on accurate prediction of enzyme commission number available at <http://www.cbrc.kaust.edu.sa/DEEPre> ([Li et al. 2018](#_ENREF_15)). |
| A machine learning- based web-server for prediction of the enzyme class. | SMV-Prot, prediction model is based on protein sequences irrespective to the similarities and available at <http://bidd2.nus.edu.sg/cgi-bin/svmprot/svmprot.cgi> ([Li et al. 2016](#_ENREF_16)). |