

# A data-driven approach utilizing a raw material database and machine learning tools to predict the disintegration time of orally fast-disintegrating tablet formulations

Navyaja Kota<sup>1</sup>, Raju Kamaraj<sup>1</sup>, S. Murugaanandam<sup>2</sup>, Mohan Bharathi<sup>1</sup>, T. Sudheer Kumar<sup>1</sup>

<sup>1</sup> Department of Pharmaceutical Regulatory Affairs, SRM College of Pharmacy, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu-603203, India

<sup>2</sup> Department of Networking and Communications, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu-603203, India

Corresponding author: Raju Kamaraj (kamarajr@srmist.edu.in)

Received 7 March 2024 ♦ Accepted 20 May 2024 ♦ Published 19 June 2024

**Citation:** Kota N, Kamaraj R, Murugaanandam S, Bharathi M, Kumar TS (2024) A data-driven approach utilizing a raw material database and machine learning tools to predict the disintegration time of orally fast-disintegrating tablet formulations. *Pharmacia* 71: 1–12. <https://doi.org/10.3897/pharmacia.71.e122507>

## Abstract

Orally fast-disintegrating tablets (OFDTs) have seen a significant increase in popularity over the past decade, becoming a rapidly expanding sector in the pharmaceutical market. The aim of the current study is to use machine learning (ML) methods to predict the disintegration time (DT) of OFDTs. In this study, we have developed seven ML models using the TPOT AutoML platform to predict the DT of OFDTs. These models include the decision tree regressor (DTR), gradient boost regressor (GBR), random forest regressor (RFR), extra tree regressor (ETR), least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), and deep learning (DL). The results indicate that ML methods are effective in predicting the DT, especially with ETR. However, after fine-tuning the deep neural network with a 10-point cross-validation scheme, the DL model showed superior performance with an NRMSE of 6.2% and an R2 of 0.79. The key factors influencing the DT of OFDTs were identified using the SHAP method.

## Keywords

Deep learning, data sets, machine learning, OFDTs, SHAP

## Introduction

Despite significant advancements in drug delivery methods, oral administration continues to be an ideal approach for administering therapeutic agents due to its precise dosage, cost-effectiveness, ability for self-medication, non-invasive nature, and convenience of administration, leading to a notable level of patient adherence. Tablets are

the most widely used type of drug, but a key limitation of these formulations is “dysphagia,” which refers to difficulties in swallowing experienced by a significant portion of the population, affecting over 50% of individuals. Consequently, individuals may fail to adhere to their prescribed drugs, leading to a high occurrence of noncompliance and inefficient treatment (Diaz et al. 2012). Recently, there has been a growing trend and acceptance of fast-disintegrating

drug delivery systems as a novel method of drug administration. These systems are preferred due to their ease of use and ability to improve patient adherence to medication. Taking conventional tablets may be challenging in certain situations, such as motion sickness, abrupt allergy responses, or coughing, and when water is not readily available, specifically for pediatric and geriatric patients. Orally disintegrating tablets (ODTs) have been developed as an alternative form of medication to address these issues. Recent developments in novel drug delivery systems (NDDS) aim to increase the effectiveness of a medicinal compound while maintaining therapeutic effectiveness, ultimately leading to improved patient adherence (Sharma 2013). Within the literature, ODTs are known by several terms, such as rapidly disintegrating, fast disintegration, fast dispersing, rapid dissolving, fast dissolve/dissolving, rapid melting, fast melting, and orodispersible tablets (Ghourichay et al. 2021). Therefore, ODTs are the optimal dosage forms due to their convenient administration, pleasant taste, and enhanced durability. They require the same equipment as traditional tablet manufacturers and have a cost-effective preparation process. OFDTs have notable drug loading capacity and dissolve quickly in the mouth without the need for water, providing compact packaging, precise dosage accuracy, and rapid disintegration in the mouth. This resulted in quick dissolution and improved absorption for a rapid pharmacological effect. OFDTs are one of the most effective dosage forms for elderly individuals, children, people with mental illnesses, and paralyzed patients (Akdaga et al. 2020). The DT of OFDTs can be affected by quality traits such as tensile strength (hardness) and porosity. Typically, when a tablet's hardness is increased, its porosity decreases, resulting in a longer DT. On the other hand, tensile strength lower than the required level may result in chipping and breakage flaws, hindering packaging and production activities and impacting both efficacy and safety. This could potentially lead to treatment failure caused by inconsistent formulations. Tablet rigidity can be improved by considering various aspects, including the selection of excipients, powder substance, compression force and rapidity, moisture level, and tablet diameter. A standard dosage form usually contains disintegrating agents, solubility and binding agents, lubricants, and additives. Component selection is determined by the specific requirements of the dosage form and assembly methods, with each feature and its amount influencing the critical quality attributes (CQAs) of a dosage form (Szlek et al. 2022). ODTs can be obtained using a variety of procedures, including lyophilization, molding, the cotton candy process, spray drying, mass extrusion, compaction, and other unique approaches. Moreover, the pharmaceutical industry finds the formulation of OFDTs through the direct compression method to be the most appealing due to its cost-effectiveness and short process. This process does not require the use of advanced machinery and technologies (Alejandro et al. 2020). However, orally disintegrating tablets (ODTs) produced through direct compression may experience a higher compression

force, resulting in increased tensile strength, which in turn may cause a longer disintegrating time (DT). The challenge with dosage forms is to have a uniform structure that allows for quick disintegration while preserving tablet hardness. Pharmaceutical experts continue to rely on traditional experimentation in product development, which is inefficient, time-consuming, and often unpredictable (Paulz et al. 2021). Meeting the critical quality attributes (CQAs) of a short DT (<180s) may pose challenges due to the complex relationship among active pharmaceutical ingredients (APIs), excipients, and the tablet manufacturing process (Szlek et al. 2022). In recent years, the use of data-based prediction technologies such as ML and DL models to streamline the conventional drug discovery and development process has increased (Yoo et al. 2022). ML is a subset of artificial intelligence (AI) that can acquire knowledge and make predictions on complex structures using extensive datasets. It has been successfully applied in pharmaceutical research across various domains. ML algorithms in the formulation field have effectively created highly precise models for forecasting the DT of ODTs and films (Hana et al. 2019). On the other hand, DL has emerged as a highly promising field of research. DL is the predominant and extensively used ML method, proven successful in drug development and repurposing through the prediction of drug-target interactions and drug evaluations. This is because DL can extract complex features from input data (Yoo et al. 2022). Additionally, DL techniques can achieve the highest accuracy in predicting the in-vitro efficacy of pharmaceutical dosage forms (Ma et al. 2020). Predicting the quality features of solid dosage forms has become a trend in various research approaches, such as explaining the disintegration process of ODTs through various ML models (Szlek et al. 2022). Forecasting breaking force of tablets and DT of tablet formulations based on ML tools (Akseli et al. 2017), computational intelligence for the prediction of DT of ODTs (Szlęk et al. 2021), DL-based models for the prediction of DT of ODTs (Yanga et al. 2019), drug properties prediction based on DL models (Yoo et al. 2022), prediction of internal tablet defects using DL Convolutional Neural Networks (Ma et al. 2020), DL in drug discovery (Askr et al. 2023), DL-based dosage predictions for radiotherapy targeting the head and neck region (Gronberg 2023), prediction of pharmacological properties of drugs using DL models (Aliper et al. 2016), quantifying the composition of amlodipine and enalapril in combination tablets with artificial neural networks (ANN) (Behei et al. 2022), prediction of DT of ODTs using ANN (Hana et al. 2018). These instances demonstrate that ML models can provide accurate predictions of CQAs, and it is possible to derive prediction rules from these models. In this context, we have developed an optimal method for evaluating the DT of OFDTs based on ML models, namely DTR, GBR, RFR, ETR, LASSO, SVM, and DL. All these models have demonstrated successful applications in the fields of pharmaceutical formulation development, manufacturing processes, and destructive analytical tests. However, effectively applying these

models requires careful consideration of data quality, model interpretability, and regulatory compliance (Loua et al. 2021). The present work aims to develop ML models capable of evaluating the DT of OFDTs. The results obtained through training, validation, and explainability are expected to enhance domain knowledge in the framework of designing formulations and optimizing process variables for manufacturing tasks.

## Methodology

### Data description

An established literature-based data model was selected for the development and validation process. The data has been refined to include only verified records, focusing on features of OFDTs such as tablet hardness, thickness, friability, and punch size. To expand our database, we conducted a literature review using the Scopus, Web of Science, and Google Scholar databases. A keyword search strategy was employed, including terms like “oral disintegrating,” “fast disintegrating,” “rapidly disintegrating,” and “oral dispersible.” The formulation should specify the total quantity of all excipients. Additionally, tablet quality attributes such as hardness, thickness, friability, punch size, and disintegration time should be included (Hana et al. 2018).

A total of 248 articles were retrieved through a database search. Out of these, only 185 research articles were selected for data extraction. Upon further manual search, 93 articles did not meet the inclusion criteria and were excluded from the study. After thorough sorting, 92 articles yielded a total of 1076 formulations. The formulation data included the name of the active pharmaceutical ingredient (API), other excipients, and process details, all of which were documented in the dataset. The final dataset consisted of the following parameters for each formulation: API name, dose, excipient name, dose (each excipient displayed in a separate column), hardness, friability, thickness, punch size, and DT. This information was then used for modeling using ML techniques (Momeni et al. 2023).

### Data enhancement and processing

According to the European Pharmacopoeia 10<sup>th</sup> edition, orodispersible tablets should breakdown within 3 minutes. Therefore, any records from the database that exceed 180 seconds have been excluded from further analysis. A correlation study was conducted to investigate the relationship between the dependent variable (DT) and various independent factors such as API, process parameters, and composition (Szlek et al. 2022).

### Workflow

The process of developing the ML model is split into three phases: data pre-processing, modeling, and model interpretation, as shown in Fig. 1.

## Pre-processing of data

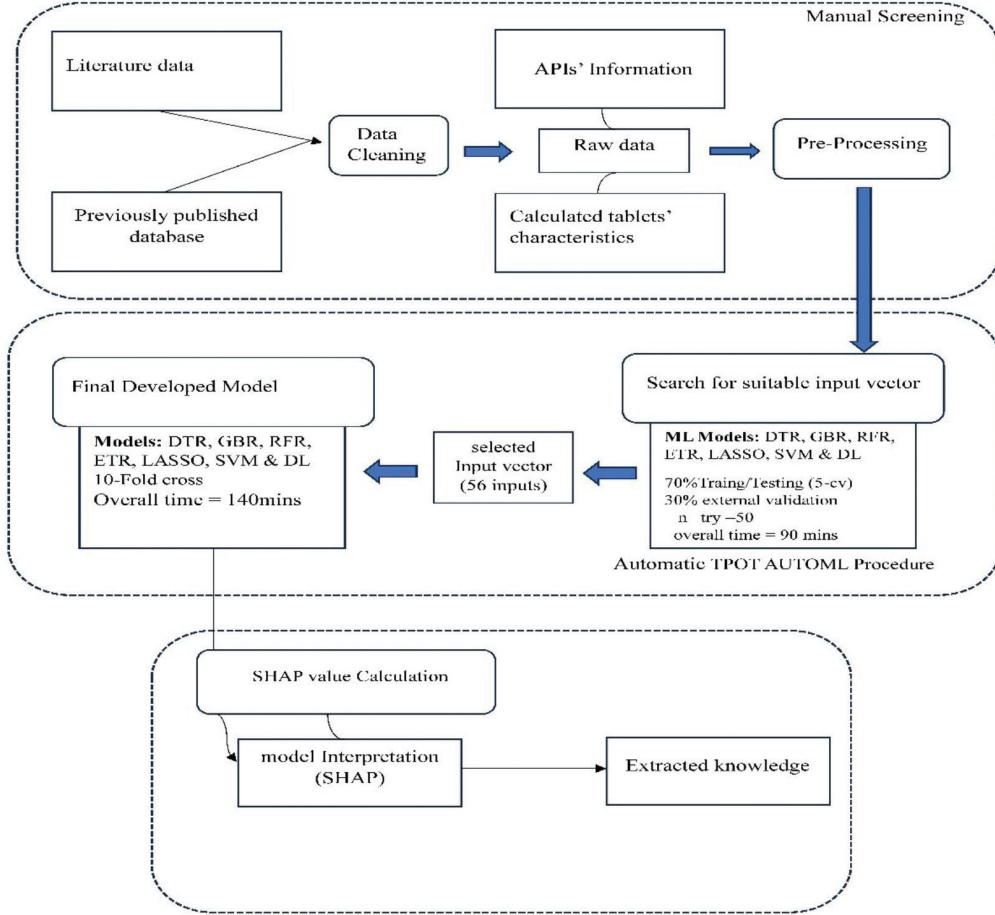
After completing data collection, the data must be processed before building predictive models to ensure the robustness and effectiveness of ML models. Several commonly employed methods, such as data cleansing, dimension reduction, imbalanced data solutions, and data splitting strategies, are necessary for data analysis. Data cleaning is performed to identify missed observations and is done by replacing data points with median or mean values. However, there are limitations to replacing missing values, as a decrease in data size may impact the accuracy of the model. Dimensionality reduction is used to eliminate the least significant features in the dataset, reducing overfitting issues and simplifying the model's complexity. Various approaches to dimensionality reduction, such as principal component analysis (PCA), high correlation filtering, and random forest feature selection, are commonly used in data processing. Imbalanced data solutions address the uneven distribution of different database classes, as using an unbalanced dataset in a prediction model can lead to poor performance. Data splitting is another crucial step in data processing, where the entire dataset is randomized and divided into three subcategories: training, validation, and testing. The training set is used to train the models; validation is for tuning hyperparameters and preventing overfitting; and the testing set is used to assess the prediction potential of unknown data. The recommended ratio for these categories is 70% for training, 20% for validation, and 10% for testing, though the ratios may vary depending on the data size. Therefore, data preprocessing and splitting strategies are essential steps before undertaking the task.

## Modeling

ML modeling tasks involve various techniques such as classification, regression trees, neural networks, and potentially many other algorithms. These models are trained using prepared databases, and their performance is evaluated using an error metric. Keeping track of different modeling methods and exploring various features can be challenging and computationally expensive. Therefore, AutoML (Automated Machine Learning) is utilized. AutoML approaches often use ensemble learning strategies, which combine several model types to produce predictions that are more reliable. In this case, TPOT AutoML employed the K-fold cross-validation technique to generate a definitive production model by selecting features based on a predefined threshold. Each fold consists of a distinct training-testing pair and a validation set, with 568 records randomly selected for training, 244 records for validation, and 348 records for testing.

## Model training

After completing the ML modeling process, it is necessary to evaluate the predictive performance of ML



**Figure 1.** Schematic representation of an applied workflow.

models to understand how well they generalize to new, unseen data. ML models are often prone to overfitting, which occurs when the model not only learns the underlying patterns in the training data but also the noise and random fluctuations that come with the data. To prevent overfitting and ensure model stability, the selected attributes indicate the significant impact they had on the model's assumptions, highlighting the opaque nature of machine learning models. A small amount of data was removed during the model evaluation using the K-fold approach. This research involves training and validating models using a five-fold cross-validation method, followed by selecting features using a Python script. The training and validation procedures were repeated five times to thoroughly cover the input database and achieve the optimal model. After selecting the final input feature vector, the model was trained using a 10-fold cross-validation procedure. Root mean square error (RMSE), normalized root mean square error (NRMSE), coefficient of determination ( $R^2$ ), mean absolute error (MAE), and mean square error (MSE) are used to measure the robustness of the models. Seven algorithms from the TPOT AUTOML platform were utilized for feature selection and final model development: DTR, GBR, RFR, ETR, LASSO, SVM, and DL.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\text{pred}_i - \text{obs}_i)^2}{n}}$$

$$NRMSE = \frac{RMSE}{\text{obs}_{\max} - \text{obs}_{\min}} (100\%)$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_{i=1}^n (\text{pred}_i - \text{obs}_i)^2}{\sum_{i=1}^n (\text{obs}_i - \text{obs})^2}$$

The variables in the equation are as follows: “obs<sub>i</sub>, pred<sub>i</sub>” represent the practical and expected values, respectively; “i” is the data record number; “n” is the total number of records; “obs<sub>max</sub>” is the highest experimental value; “obs<sub>min</sub>” is the least observed value; “R<sup>2</sup>” is the coefficient of determination; “SS<sub>res</sub>” is the sum of squares of the residual errors; “SS<sub>tot</sub>” is the total sum of the errors; and “obs” is the arithmetic mean of observed values (Szlek et al. 2022).

## Proposed ML method

The accuracy of ML results cannot be improved simply by fitting data into models. As the data becomes larger and more complex, better data handling techniques such as DTR, GBR, random RFR, ETR, LASSO, SVM, and DL become necessary to handle it.



## DTR

DTR is a versatile algorithm used for classification and regression tasks simultaneously. It operates on the concept of breaking down complex problems into simpler, more manageable subproblems, making it an excellent choice for various applications. Decision trees (DT) have a hierarchical structure, with conditions applied from the tree's root to its leaves. This structure allows for a step-by-step decision-making process. One of the key strengths of DT is its transparent and interpretable structure. The rules generated by DT are easy to understand. Once trained on a dataset, DT can produce logical rules that can be applied to new, unseen data by recursively dividing them into subgroups based on the conditions learned during the training phase.

## GBR

GBR generates a series of decision trees, with each tree addressing the errors of the previous one. The model is generated iteratively, with each iteration adding a new decision tree to the ensemble and focusing on the errors or residuals of the combined model from prior iterations. The loss function is a crucial component in GBR as it determines the variation between the predicted and actual values of the desired variables. The algorithm minimizes this loss function during each iteration, ensuring that the model is continually improving. MSE is a commonly used loss function in GBR, where the average squared difference between the expected and actual values is calculated. Overall, GBR is a powerful algorithm for regression tasks and is widely used in practice due to its flexibility, high predictive accuracy, and ability to handle complex relationships in data (Ghazwani et al. 2023).

$$f(x) = \sum_{m=1}^M \beta_m h_m(x)$$

## RFR

RFR is a machine-learning-based regression algorithm. It is built on bagging and random subspace algorithms. Due to its versatility, capability to handle uncertain data, and suitability for high-dimensional feature spaces (with many predictors), RFR is widely respected. In recent years, RFR has emerged as the most advantageous general-purpose algorithm. The “divide and conquer” approach, which involves bootstrapping data subsets, building decision trees on each subset, and then aggregating these results, best characterizes.

$$\hat{f}_{RFR}^C(x) = \frac{1}{C} \sum_{i=1}^C T_i(x)$$

The RFR employs a vector input variable  $x$  to generate an output. This is achieved by merging the predictions of the  $C$  decision trees.  $T_i(x)$  indicates a regression tree created from a subset of input variables and bootstrapped samples (Borup et al. 2023).

## ETR

ETR is an enhanced method for addressing generalization (overfitting) concerns associated with random forest (RF). This approach is a recent advancement in the field of ML and can be viewed as an extension of the widely used RF. Its purpose is to minimize the risk of overfitting. Similar to RF, ETR trains each base estimator using a random subset of features. It does not select a feature and its corresponding value for use in node splitting (Hameed et al. 2021).

## LASSO

LASSO is a linear regression method that reduces the total sum of squares of residuals and the sum of the absolute values of the regression coefficients. The regression coefficient can be obtained using the given equation.

$$\sum_{i=1}^m (y^{(i)} - x^{(i)}b)^2 + \lambda \sum_{j=1}^n |b_j|$$

Where  $m$  represents the number of samples,  $n$  represents the number of  $x$  variables -  $y^{(i)}$  and  $x^{(i)} \in \mathbb{R}^{1 \times m}$  are the  $y$  and  $x$  values in the  $i$ th sample, respectively;  $b_j$  is the  $j$ th regression coefficient; and  $\lambda$  is the hyperparameter.  $b$  is denoted by the following formula:

$$b = (b_1 \ b_2 \ \dots \ b_n)^T$$

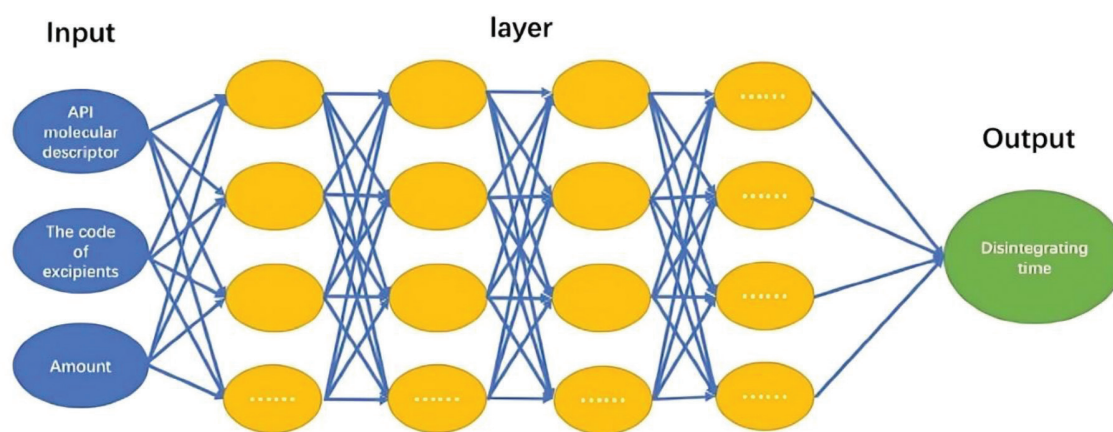
In LASSO, the regression coefficient,  $b_j$ , can be reduced to zero, leading to the removal of the corresponding  $x$ . The study considered a range of values for  $\lambda$  from 2–15 to 2–14 ..., 2–2, and 2–1 to find the value that maximizes the coefficient of determination,  $r^2$ , by 5-fold cross-validation. Scikit-learn was employed to estimate the LASSO (Kaneko 2021).

## SVM

SVM is the most commonly used method in ML for classification, regression, and other tasks. SVM operates in high- or infinite-dimensional space and constructs a hyperplane or multiple hyperplanes. The hyperplane that maximizes the distance from the closest training data points in each class achieves significant separation. A larger margin typically results in a lower generalization error for the classifier. It is effective in high-dimensional spaces and can exhibit different behaviors based on mathematical functions like the kernel. SVM classifiers utilize various types of functions, such as linear, polynomial, radial basis function (RBF), and sigmoid, as kernel functions. However, if the dataset contains a higher level of noise, such as overlapping target classes, the performance of SVM is compromised (Gaye et al. 2021; Pérez and Bajorath et al. 2022).

## DL

DL is primarily used as a neural network, as shown in Fig. 2. DL can automatically extract features and transform



**Figure 2.** The network architecture of DNN.

basic representations into more complex abstraction layers without needing a separate feature extractor. DL is more sensitive to small and specific modifications in complex networks, leading to higher accuracy than typical ML techniques. DL algorithms have shown superior performance compared to other ML methods in effectively forecasting the in vitro performance of pharmaceutical formulations. Deep Neural Networks (DNN) have been utilized in pharmaceutical research, particularly in drug design, drug-induced liver toxicity, and virtual screening. Deep learning can create sophisticated and complex systems that represent various objects using chemical descriptors. This can greatly aid in the development of drugs and their prediction (Roggo et al. 2020).

## Model interpretation

As ML models are inherently black boxes, efforts have been made to shed light on their prediction techniques. In our study, illustrated in Fig. 1, we used the Lundberg et al. SHAPLEY additive explanation (SHAP) approach to explain the relationship between input and output variables. The SHAP method is primarily based on cooperative game theory and has been applied in various domains, including the pharmaceutical industry. The concept of Shapley values is used to assess the contribution of each participant or individual to the overall team effort or outcome. In machine learning, Shapley values have been adapted to explain the impact of each feature in a predictive model. SHAP values offer a method to distribute the model's prediction to each feature in a fair and consistent manner. The mathematical formula for SHAP values is provided below.

$$\phi_j(\text{val}) = \sum_{S \subseteq \{1, \dots, p\} \setminus \{j\}} \frac{|S|!(p - |S| - 1)!}{p!} (\text{val}(S \cup \{j\}) - \text{val}(S))$$

Where “S” is a subset of features in the model, “x” is the vector of feature values to be explained, and “p” is the number of features. The prediction “Valx(S)” is the result of estimating feature values in set “S” (Rozemberczki et al. 2022).

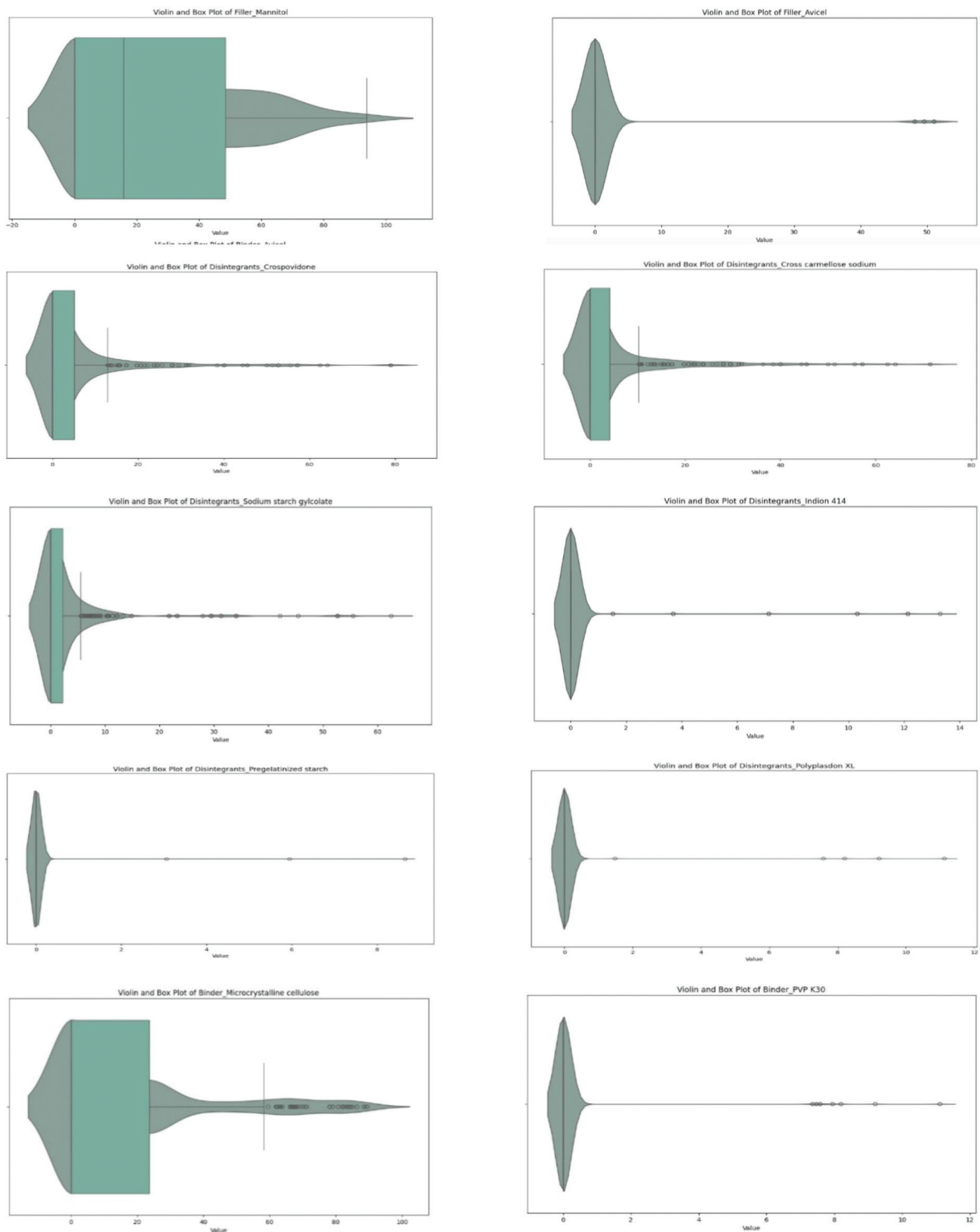
The Shapley value calculation method adheres to the axioms of efficiency, symmetry, dummy, and additivity, thus explaining the process for developing predictions. Random samples are used to replace the values of each attribute to assess their importance and impact. Computing the Shapley value can be computationally intensive due to the numerous potential coalitions of feature values that must be considered. Coalitions are carefully chosen to reduce repetitions, leading to decreased calculation time. However, this approach also leads to an increase in the variation of the Shapley value. The k-means method was utilized to reduce the number of repetitions needed to represent each feature's impact. The k-means algorithm was set up with 12 centroids, each corresponding to a feature data domain in a cluster. A comprehensive SHAP matrix can be created by grouping the data domain of each characteristic. Displaying this matrix facilitates the understanding of the model's predictions (Szlek et al. 2022).

## Results

### Database

The pre-processed database contained 92 direct compression OFDTs (new data entries), including 28 unique APIs and 50 variable coding compositions (excipients were topologically coded). There were 5 variables encoding formulation dimensions such as thickness [mm], hardness [N], friability [%], punch size [mm], and DT [s]. Descriptive statistics (Table 1) show that the variables did not follow a normal distribution and the formulations were significantly positively skewed (right-skewed distribution), as shown in Fig. 3. The database was divided using a 10-fold cross-validation method that was calibrated to ensure input variables were classified fairly across the splits.

The raw data and curated data are available at (Raw database) <https://doi.org/10.6084/m9.figshare.25880377>. (Curated database) <https://doi.org/10.6084/m9.figshare.25880560>. (accessed on 22 May 2024)



**Figure 3.** Box and violin plots of specific features from the database. The graph illustrates the interquartile range (IQR) using boxes, which include the first quartile (Q1), median (horizontal line), and third quartile (Q3). The lower whisker is calculated as  $Q1 - 1.5 \cdot IQR$  and the upper whisker as  $Q3 + 1.5 \cdot IQR$ . These graphs demonstrate the dispersion of numerical data through the kernel density function.

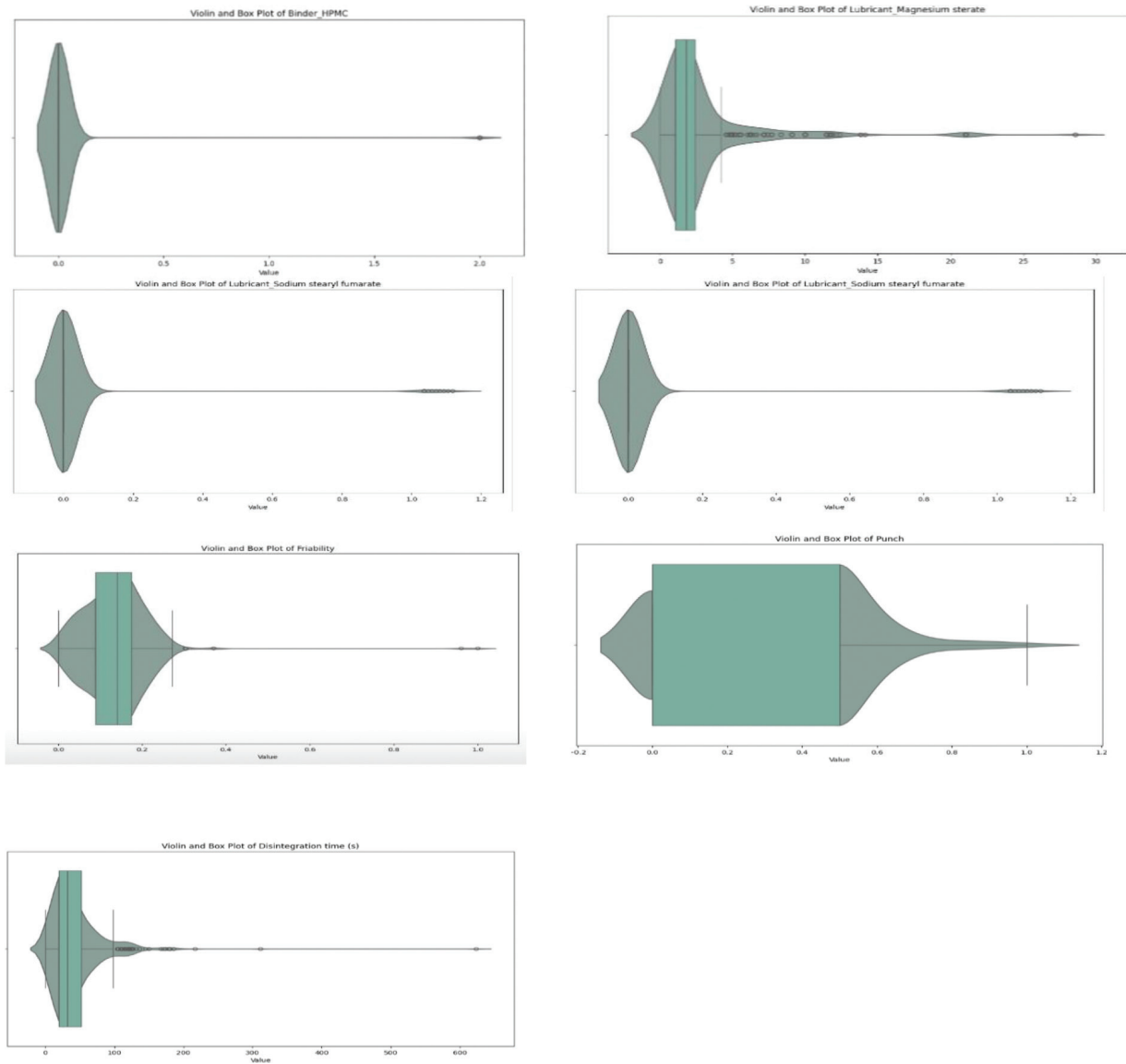


Figure 3. Continued.

Table 1. Descriptive statistics of the dataset.

Feature	Count	Min	Mean	Std	25%	50%	75%	Max
Thickness (nm)	690	2.9	1.19	0.82	2.4	3.14	3.6	6.5
Hardness	690	3.42	0.85	0.17	3.1	3.48	4	7.98
Friability	690	0.55	0.27	0.1	0.4	0.57	0.68	3.45
Punch	690	5.6	4.1	0	0	8	8	16
Disintegration time (s)	690	41.62	40.37	0.47	20	32	51.9	623
Filler Avicel [%]	690	0.86	6.48	0	0	0	0	51.02
Filler Mannitol [%]	690	25.26	27.5	0	0	15.79	48.52	93.72
Binder Avicel [%]	690	2.71	9.93	0	0	0	0	58.42
Binder HPMC [%]	690	0.02	0.19	0	0	0	0	2
Binder Microcrystalline cellulose [%]	690	15.36	23.94	0	0	0	23.65	89.29
Binder_PVP K30 [%]	690	0.09	0.85	0	0	0	0	11.11
Disintegrants Crospovidone [%]	690	5.05	11.32	0	0	0	5.15	78.95
Disintegrants Cross carmellose sodium [%]	690	4.44	10.26	0	0	0	4.1	71.43
Disintegrants Indion 414 [%]	690	0.12	1.08	0	0	0	0	13.3
Disintegrants Polyplasdon XL [%]	690	0.05	0.7	0	0	0	0	11.11
Disintegrants Pregelatinized starch [%]	690	0.03	0.42	0	0	0	0	8.65
Disintegrants Sodium starch glycolate [%]	690	2.52	7.24	0	0	0	2.22	62.5
Lubricant Magnesium stearate [%]	690	2.8	3.64	0	1.05	1.81	2.42	28.57
Lubricant Sodium stearyl fumarate [%]	690	0.02	0.15	0	0	0	0	1.12



## Choosing features and developing the final model

The selection of features and development of the final model were conducted using an automated approach using the TPOT AUTOML method. The dimensions of the TPOT AUTOML are provided in Table 2. This table displays the accuracy of the models that were developed. The values of RMSE, NRMSE,  $R^2$ , MAE, and MSE are also analyzed to assess the accuracy and precision of the model's output.

**Table 2.** Robustness of the TPOTAUTOML developed models.

ML Techniques	RMSE (s)	NRMSE (%)	$R^2$	MAE	MSE
DTR	33.18	10.0	0.28	13.28	1156.34
GBR	27.1	9.0	0.5	12.81	863.83
RFR	25.98	7.0	0.57	12.39	745.91
ETR	23.37	7.0	0.65	9.55	608.58
LASSO	26.38	8.0	0.55	13.38	761.85
SVM	30.04	9.0	0.4	14.18	970.99
DL (5-F-CV)	30.6	6.89	0.61	17.1	940.34
DL (10-F-CV)	27.9	6.29	0.79	14.8	782.66

As expected, ML techniques, especially ETR, have proven to be the best pipeline in the TPOT AutoML analysis for the curated dataset. However, it is clear that the deep learning model, when hyper-tuned, performed on par with its ML counterparts, achieving great  $R^2$  values and NRMSE percentages. After the initial evaluation of the deep learning models trained using a five-fold cross-validation scheme, it was found that when the DL was trained with three hidden layers of 100 neurons each, 56 input neurons, and one output neuron, using a tanh activation function with 2200 epoch values combined with a 10-fold cross-validation scheme, the model accuracy significantly improved. This improvement was reflected in the NRMSE and  $R^2$  values, as shown in Fig. 4.

## Feature selection of input variables based on scaled importance

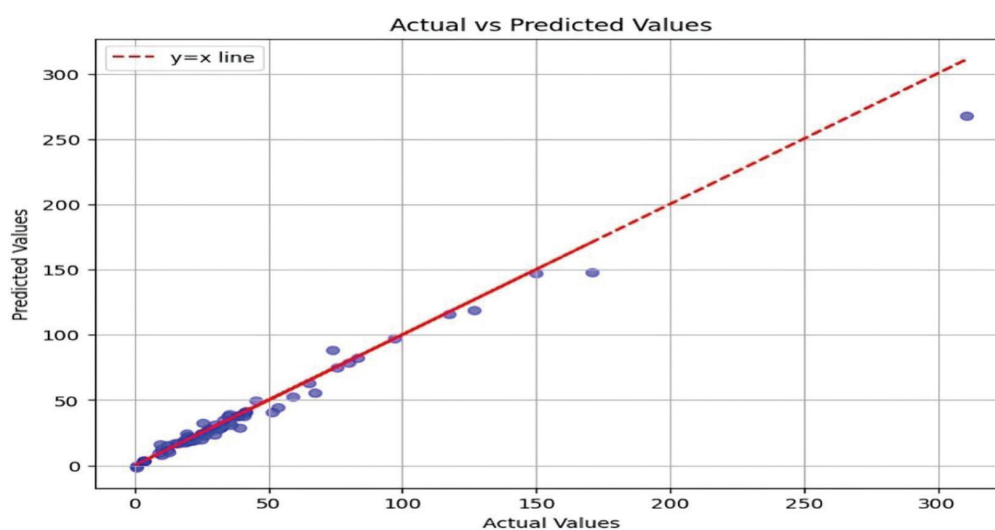
The input variables were categorized into two main groups: composition and manufacturing parameters. Features below the variable importance level were eliminated, except for those in the composition section. The final input vector consisted of 18 inputs. (Table 3) displays the chosen features and their scaled significance.

Table 3 shows that the quantities of disintegrants (Crospovidone and croscarmellose sodium) and binder (Microcrystalline cellulose) have varying levels of significance. The data suggests that the number of disintegrants will have the most significant impact on the predicted outcomes. The lower significance of components like lubricants and fillers may be due to the positive skewness of the variable distribution.

## Model interpretation

The SHAP summary graph in Fig. 5 illustrates the influence of features on the model's predicted output. The colored line shows actual feature values and their impact on a prediction along the x-axis. A SHAP summary plot helps identify overall impacts and underlying assumptions.

A higher disintegration time (DT) is predicted with a greater concentration of disintegrants such as crospovidone and croscarmellose sodium. Conversely, when it comes to fillers like Mannitol and Avicel, a higher quantity of Mannitol and Avicel leads to lower DT. For binders like Microcrystalline Cellulose (MCC), Hydroxypropylmethylcellulose (HPMC), and Polyvinyl Pyrrolidone (PVPK30), two distinct effects are observed. A higher amount of HPMC and PVPK30 results in higher DT, while MCC decreases DT at higher concentrations. Lubricants like magnesium stearate (MgSt) and sodium stearyl fumarate (SSF) also play a role. SSF tends to increase DT, likely due to its hydrophilic properties. On the other hand, a higher amount of MgSt, which is more lipophilic, lowers DT due to the occlusion effect (Szlek et al. 2022).



**Figure 4.** Scatter plot between actual and predictive values for the disintegration time of the DL model.

**Table 3.** Selected input variables for the best predictive ML models.

Feature	Feature type	Scaled feature importance
Crospovidone [%]	Disintegrant, Composition	1
Microcrystalline cellulose [%]	Binder, Composition	0.744682848
Sodium starch glycolate [%]	Disintegrant, Composition	0.488169243
Friability	Manufacturing Parameter	0.251528875
Avicel [%]	Filler, Composition	0.104084555
Indion 414 [%]	Disintegrant, Composition	0.089409042
Thickness(mm)	Manufacturing Parameter	0.087242712
Pregelatinized starch [%]	Disintegrant, Composition	0.081137918
Polyplasdon XL [%]	Disintegrant, Composition	0.074485462
HPMC [%]	Binder, Composition	0.045193837
Cross carmellose sodium [%]	Disintegrant, composition	0.039986272
Avicel [%]	Binder, Composition	0.016576396
Sodium stearyl fumarate [%]	Lubricant, Composition	0.005441542
Mannitol [%]	Filler, Composition	0.00258793
Punch	Manufacturing Parameter	0.002117971
Hardness	Manufacturing Parameter	0.001023564
Magnesium stearate [%]	Lubricant, Composition	0.000947287
PVP K30 [%]	Binder, Composition	0.000403805

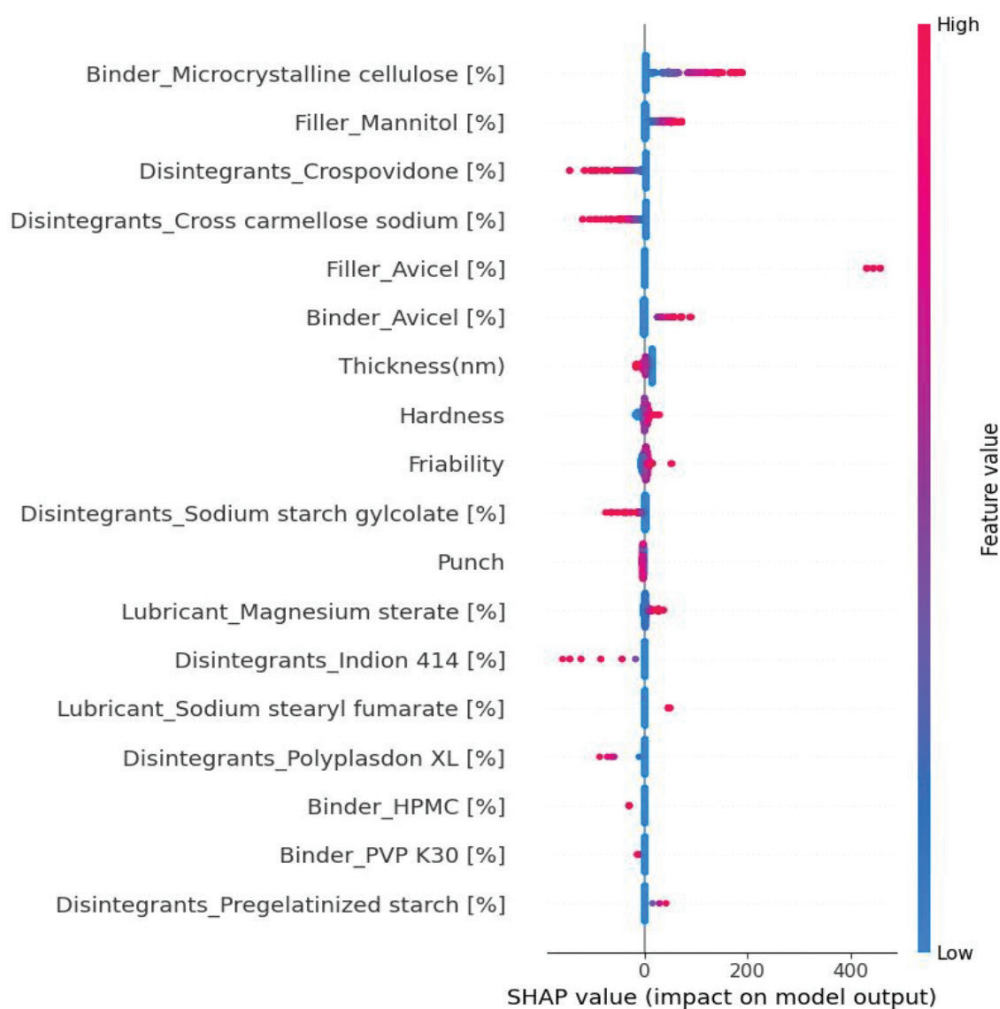
## Discussion

OFDTs have experienced a significant increase in demand over the past decade, leading to rapid growth in the pharmaceutical sector. Oral drug delivery remains the preferred method for administering many drugs. Advances in technology have inspired researchers to develop OFDTs that improve patient compliance and convenience. These tablets disintegrate upon administration without the need for water, making them popular and useful for various patient populations, particularly pediatric and geriatric individuals who may have difficulty swallowing traditional tablets and capsules (Parkash et al. 2011). Developing effective ODTs is a challenging task, even for experienced pharmaceutical professionals. Various factors, including materials, manufacturing processes, analytics, and regulatory requirements, must be carefully considered to ensure the products meet standards for effectiveness, safety, stability, and processability (Loua et al. 2021). The demand for high-quality OFDT formulations utilizing innovative disintegrants and efficient manufacturing techniques has been on the rise in recent years. Three common processes used in OFDT manufacturing are freeze-drying, tablet molding, and tablet compression. Direct compression is the preferred method due to its effectiveness and simplicity. ODTs formulated through direct compression typically contain filler, binder, disintegrant, lubricant, and solubilizer. Developing a prototype of ODT formulations is essential to

minimize disintegration time while maintaining high tablet quality (Hana et al. 2018).

Pharmaceutical formulation manufacturing currently relies on the trial-and-error method, which is both ineffective and time-consuming. In recent years, ML has emerged as a solution that can generate data-driven forecasts using existing experimental data, opening up significant possibilities for creating optimal formulations. A well-established ML algorithm can greatly speed up the development process, optimize formulations, save costs, and maintain product consistency (Yanga et al. 2019). DL has become more prevalent in pharmaceutical research in the last five years. DL and ML have revolutionized the world's perspective. The DT is a CQA for OFDTs that can be optimized. Therefore, predicting the DT of OFDTs is a crucial step in pharmaceutical development. The DT can be influenced by several factors, with the type and concentration of disintegrant used in the formulation being significant elements that affect tablet disintegration. The compaction pressure applied to the tablet also plays a crucial role in its disintegration. Typically, increasing the compaction pressure will result in a longer DT. This is because higher compaction pressures lead to the formation of stronger interparticle bonds, which take more time to be disrupted in a disintegration test. Increased compaction pressure reduces tablet porosity, hindering liquid penetration and delaying tablet disintegration. However, if tablets have high porosity, the effectiveness of disintegrants may decrease because the swelling pressure of the disintegrant is partially reduced by being accommodated in the large empty spaces. Essentially, if the compaction pressure is either very high or excessively low, it may result in prolonged disintegration. Interestingly, it has been observed that increasing the compaction pressure accelerates the disintegration of tablets containing crospovidone. This is due to its enhanced strain recovery at elevated compaction pressures (Zheng et al. 2022).

In the present study, as mentioned in the methodology section, the developed models were analyzed to determine which produced better outcomes. The outcomes displayed in Table 2 are based on the 5-fold cross-validation scheme. The ML technique, especially ETR, showed the best results in the TPOT AUTOML analysis. However, after hypertuning, the DL model resulted in a higher  $R^2$  and NRMSE%. Following the TPOT AUTOML output analysis, we proceeded with the final model development using a 10-fold cross-validation scheme. The DL model was trained over 2200 epochs, yielding better results than previous models. Additionally, critical parameters affecting the DT were plotted in Fig. 4. A thorough analysis was conducted to understand the underlying principles of the final model's forecasting by utilizing Shapley values. The findings not only provide insights into DT prediction but also support the effectiveness of autoML-based approaches in addressing the challenges of complex pharmaceutical tasks. However, the results showed that DL achieved significant outcomes compared to other ML models. Therefore, our predictions were consistent with previous studies (Akseli et al. 2017; Hana et al. 2018; Szłęk et al. 2021; Szlek et al. 2022; Mehri et al. 2024).



**Figure 5.** SHAP dependence diagram for the ML models for the top 20 attributes.

## Conclusion

OFDTs are a promising method to achieve rapid pharmacological action and offer advantages over traditional dosage forms already on the market. The conventional method of formulation development, based on trial and error, is tedious and demanding. In contrast, a ML-driven development technique accelerates the process by allowing scientists to efficiently produce accurate predictions. ML and DL models were effectively created in this study to forecast the DT of OFDTs. Although ML models are often deemed inscrutable black boxes, with the use of theoretical approaches such as Shapley additive explanations, we can gain an estimated understanding of what's happening

inside the black box. The study outcomes exhibited that the proposed ML models could precisely predict the DT of OFDTs, with DL showing better performance and lower complexity compared to other established models. Therefore, DL could also be applied to more fields in pharmaceutical research. The anticipated benefits of DL include a substantial reduction in the duration of therapeutic product development and a decrease in the quantity of materials required. Furthermore, the interdisciplinary fusion of pharmaceuticals and AI has the potential to transform pharmaceutical research from experience-based studies to data-driven approaches. Various ML techniques will be investigated in the future to forecast optimal formulations more effectively.

## References

- Akdaga Y, Gulsuna T, Izata N, Cetin M, Onera L, Sahin S (2020) Characterization and comparison of deferasirox fast disintegrating tablets prepared by direct compression and lyophilization method. *Journal of Drug Delivery Science and Technology* 57: 1–8. <https://doi.org/10.1016/j.jddst.2020.101760>
- Akseli I, Xie J, Schultz L, Ladyzhynsky N, Bramante T, He X, Deanne R, Horspool KR, Schwabe R (2017) A Practical Framework Toward Prediction of Breaking Force and Disintegration of Tablet Formulations Using Machine Learning Tools. *Journal of Pharmaceutical Sciences* 106: 234–247. <https://doi.org/10.1016/j.xphs.2016.08.026>
- Alejandro B, Guillermo T, Angeles PM (2020) Formulation and Evaluation of Loperamide HCl Oro Dispersible Tablets. *Pharmaceutics* 13: 1–24. <https://doi.org/10.3390/ph13050100>

- Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A (2016) Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptional Data. *Molecular Pharmaceutics* 13: 2524–2530. <https://doi.org/10.1021/acs.molpharmaceut.6b00248>
- Askr H, Elgelda E, Ella HA, Elshaie YAAM, Gomaa MM, Hassanien AE (2023) Deep learning in drug discovery: an integrative review and future challenges. *Artificial Intelligence Reviews* 56: 5975–6037. <https://doi.org/10.1007/s10462-022-10306-1>
- Behei N, Tryhubchak O, Pryymak B (2022) Development of amlodipine and enalapril combined tablets based on quality by design and artificial neural network for confirming of qualitative composition. *Pharmacia* 69: 779–789. <https://doi.org/10.3897/pharmacia.69.e86876>
- Borup D, Christensen BJ, Muhlbach NS, Nielsen MS (2023) Targeting predictors in Random Forest Regression. *International Journal of Forecasting* 39: 841–868. <https://doi.org/10.1016/j.ijforecast.2022.02.010>
- Diaz JEA, Montoya EG, Negre JMS, Lozano PP, Minarro M, Tico JR (2012) Predicting orally disintegrating tablets formulations of ibuprofen tablets: An application of the new SeDeM-ODT expert system. *European Journal of Pharmaceutics and Biopharmaceutics* 80: 638–648. <https://doi.org/10.1016/j.ejpb.2011.12.012>
- Gaye B, Zhang D, Wulamu A (2021) Improvement of Support Vector Machine Algorithm in Big Data Background. *Mathematical Problems in Engineering* 2021: 1–9. <https://doi.org/10.1155/2021/5594899>
- Ghazwani M, Begum MY (2023) Computational intelligence modelling of hyoscine drug solubility and solvent density in supercritical processing: gradient boosting, extra trees, and random forest models. *Scientific Reports* 13: 1–12. <https://doi.org/10.1038/s41598-023-37232-8>
- Ghourichay MP, Kiaie SH, Nokhodchi A, Javadzadeh Y (2021) Formulation and Quality Control of Orally Disintegrating Tablets (ODTs): Recent Advances and Perspectives. *Bio Med Research International* 2021: 1–12. <https://doi.org/10.1155/2021/6618934>
- Gronberg MP, Beadle BM, Garden AS, Skinner H (2023) Deep Learning-Based Dose Prediction for Automated, Individualized Quality Assurance of Head and Neck Radiation Therapy Plans. *Practical Radiation Oncology* 13: 282–291. <https://doi.org/10.1016/j.prro.2022.12.003>
- Hameed MM, AlOmar MK, Khaleel F, Ansari NA (2021) An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical Applications in the Hydraulic Sector and Future Research Directions. *Mathematical Problems in Engineering* 10: 1–19. <https://doi.org/10.1155/2021/7001710>
- Hana R, Xiong H, Yea Z, Yang Y, Huang T, Jing Q, Lua J, Pan H, Ren F, Ouyang D (2019) Predicting physical stability of solid dispersions by machine learning techniques. *Journal of Controlled Release* 311: 16–25. <https://doi.org/10.1016/j.jconrel.2019.08.030>
- Hana R, Yanga Y, Lib X, Ouyang D (2018) Predicting oral disintegrating tablet formulations by neural network techniques. *Asian Journal of Pharmaceutical Sciences* 13: 336–342. <https://doi.org/10.1016/j.ajps.2018.01.003>
- Jiang J, Ma X, Ouyang D, Williams RO (2022) Emerging Artificial Intelligence (AI) Technologies Used in the Development of Solid Dosage Forms. *Pharmaceutics* 14: 1–26. <https://doi.org/10.3390/pharmaceutics14112257>
- Kaneko H (2021) Examining variable selection methods for the predictive performance of regression models and the proportion of selected variables and selected random variables. *Heliyon* 7: 1–12. <https://doi.org/10.1016/j.heliyon.2021.07356>
- Loua H, Lian B, Hageman MJ (2021) Applications of Machine Learning in Solid Oral Dosage Form Development. *Journal of Pharmaceutical Sciences* 110: 3150–3165. <https://doi.org/10.1016/j.xphs.2021.04.013>
- Ma X, Kittikunakorn N, Sorman B, Xi H, Chen A, Marsh M, Mongeau A, Piche N, Williams RO, Skomski D (2020) Application of Deep Learning Convolutional Neural Networks for Internal Tablet Defect Detection: High Accuracy, Throughput, and Adaptability. *Journal of Pharmaceutical Sciences* 109: 1547–1557. <https://doi.org/10.1016/j.xphs.2020.01.014>
- Momeni M, Rakhshani S, Abbaspour M, Alizadeh F, Sheikhi N, Zadeh FG, Habibi Z, Tabesh H (2023) Dataset development of pre-formulation tests on fast disintegrating tablets (FDT): data aggregation. *BMC Research Notes* 16: 1–5. <https://doi.org/10.1186/s13104-023-06416-w>
- Parkash V, Maan S, Deepika, Yadav SK, Hemlata, Jogpal V (2011) Fast disintegrating tablets: Opportunity in drug delivery system. *Journal of Advanced Pharmaceutical Technology & Research* 2: 223–235. <https://doi.org/10.4103/2231-4040.90877>
- Paulz D, Sanapz G, Shenoy S, Kalyane D, Kalia K, Tekade RK (2021) Artificial intelligence in drug discovery and development. *Drug Discovery Today* 26: 80–93. <https://doi.org/10.1016/j.drudis.2020.10.010>
- Pérez RR, Bajorath J (2022) Evolution of Support Vector Machine and Regression Modelling in Chemoinformatic and Drug Discovery. *Journal of Computer-Aided Molecular Design* 36: 355–362. <https://doi.org/10.1007/s10822-022-00442-9>
- Roggo Y, Jelsch M, Heger P, Ensslin S, Krumme M (2020) Deep learning for continuous manufacturing of pharmaceutical solid dosage form. *European journal of Pharmaceutics and biopharmaceutics* 153: 95–105. <https://doi.org/10.1016/j.ejpb.2020.06.002>
- Rozemberczki B, Watson L, Bayer P, Yang HT, Kiss O, Nilsson S, Sarkar R (2022) The Shapley Value in Machine Learning. *Arxiv* 1–8. <https://doi.org/10.24963/ijcai.2022/778>
- Sharma D (2013) Formulation Development and Evaluation of Fast Disintegrating Tablets of Salbutamol Sulphate for Respiratory Disorders. *ISRN Pharmaceutics* 2013: 1–8. <https://doi.org/10.1155/2013/674507>
- Szlek J, Khalid MH, Paclawski A, Czub N, Mendyk A (2022) Puzzle out Machine Learning Model-Explaining Disintegration Process in ODTs. *Pharmaceutics* 14: 1–23. <https://doi.org/10.3390/pharmaceutics14040859>
- Szłęk J, Paclawski A, Czub N, Mendyk A (2021) Computational intelligence model of orally disintegrating tablets – an attempt to explain disintegration process. *Proceedings* 68: 1–6. <https://doi.org/10.3390/ASEC2021-11163>
- Yanga Y, Yea Z, Sua Y, Zhao Q, Lib X, Ouyang D (2019) Deep learning for in vitro prediction of pharmaceutical formulations. *Acta Pharmaceutica Sinica B* 9: 177–185. <https://doi.org/10.1016/j.apsb.2018.09.010>
- Yoo S, Kim J, Choi GJ (2022) Drug Properties Prediction Based on Deep Learning. *Pharmaceutics* 14: 1–11. <https://doi.org/10.3390/pharmaceutics14020467>
- Zheng AY, Heng PWS, Chan LW (2022) Tablet Disintegrability: Sensitivity of Superdisintegrants to Temperature and Compaction Pressure. *Pharmaceutics* 14: 1–15. <https://doi.org/10.3390/pharmaceutics14122725>
- Mehri M, Marziyeh A, Saleh R, Amin M, Hamed T (2024) A prediction model based on artificial intelligence techniques for disintegration time and hardness of fast disintegrating tablets in pre-formulation tests. *BMC Medical Informatics and Decision Making* 24: 1–12. <https://doi.org/10.1186/s12911-024-02485-4>