PREDICTION OF SUITABLE NANO DRUG DELIVERY FOR CANCER TREATMENT THROUGH AI

K.Jamberi[1], Pankaj Ramesh Gavit[2], Santosh Kumar Nathsharma[3], K.Kowsalya[4], M D Kitukale[5], Manmohan Singhal[6], Manideep Karukuri[7]

Article History: Received: 02.10.2022 **Revised:** 23.12.2022 **Accepted:** 15.03.2023

Abstract

Nano-Tumor Database, which is recently released utilizing information, produced from a Physiologically Based Model (PBM) mode that contains 376 datasets, this study examined tumor models, the effects of NP physicochemical properties, and cancer types on NP tumor delivery efficiency. Outperforming all other machine learning techniques, the deep neural network model accurately predicted the effectiveness of various NPs in treating various tumors, consisting of Linear Regression, Support Vector Machine, Bagged model, and Random Forest techniques. To increase tumor delivery efficiency and to improve the design of cancer nanomedicine, this study offers a quantitative model. Our comprehension of the reasons for low NP tumor delivery efficiency is enhanced by these findings. This research shows that it is possible to study cancer nanomedicine by combining Artificial Intelligence with PBM modeling techniques.

Keywords; Nano Medicine; Physicochemical; Machine Learning; Deep Neural Network; Quantitative Model

[1]Assistant Professor, Department of Computer Science, S.A. College of Arts & Science, Veeraraghavapuram,chennai-77.
[2]Assistant Professor, Department of Chemistry, Sant Dnyaneshwar Mahavidyalaya, Soegaon Aurangabad, Maharashtra 431120,India
[3]Lecturer, Department of Chemistry, Christ College, Cuttack
Odisha, 753008, India,
[4]Assistant Professor, Department of ECE, Hindusthan Institute of Technology, Coimbatore-32, Tamilnadu,India,
[5]Professor ,P Wadhwani College of Pharmacy , Yavatmal ,India
[6]Associate Professor, School of Pharmaceutical and Population Health Informatics, Faculty of Pharmacy, DIT University, Dehradun, Uttarakhand 248009. India
[7]University of Texas at Arlington, MSBA Graduate, DALLAS, TEXAS
United States,

Email: [1]kjamberi111@gmail.com, [2]pankajgavit@rediffmail.com, [3]sknathsharma5388@gmail.com, [4]Kowsalya.k@hit.edu.in, [5]kitukalemanisha5@gmail.com, [6]manmohan.singhal@dituniversity.edu.in, [7]manideepkarukoori@gmail.com

DOI: 10.31838/ecb/2023.12.s3.151

1. Introduction

Data science Robotics, modeling, and Automation advancements over more than the past ten years have had a significant impact on or will soon have an impact on, most fields of science and technology. Robotics and automation, which are also the driving forces behind the phenomenal growth in omics technologies, have made it possible to synthesize and characterize materials much faster than with traditional experiments[1]. The complexity of the information gathered has also increased dramatically as a result of these developments, in addition to the number of materials that can be studied and synthesized. For instance, from omics technologies and high-content imaging[2-3]. As a result, the amount of data being collected has exponentially increased. With an urgent need for processing for extracting useful scientific meaning and computational methods, the formation of "data lakes" had various materials of extensive collections from frequently highly multidimensional data sets [4].

Numerous compartment-based mathematical models were the focus of the review as well as forecasting nanomaterial properties and the use of QSAR methods for modeling [5]. Shatkin published a second succinct article on the future of nanosafety while this one was being reviewed [6]. International investments in nanosafety have several definite advantages, including interdisciplinary collaborations [7]. Listed the present urgent problems: throughout the product life cycle, health risks and assessing environmental and health risks, future safety assessments of more advanced materials are required; creation of trustworthy and pertinent new techniques for assessing safety with less mammalian testing. To advance the idea of safety by design, she pushed for the creation of screening techniques that are quicker and more effective.

Numerous unit operations are required for the modeling of the biological properties of nanomaterials.[8]. To create a data set for ML model training, materials are synthesized and tested in pertinent biological assays [9-11]. Descriptors are mathematical symbols that must be used to represent the physicochemical characteristics of nanomaterials. From a list of descriptors, suitable relevant features are chosen, and created a predictive model of the desired property using ML techniques [12]. Unrelated materials that were not used to create the models must be predicted to validate the models. Alternatively, omitting a material or materials from the model during cross-validation techniques. The model produced using the leftover components predicts the material that was withheld [13].

2. Related Works

Several hundred datasets from numerous different kinds of NPs in various tumor types were used to calibrate this model. Using this model, it was possible to predict how well different NPs would perform when administered to tumor-bearing mice at various points after intravenous injection [14–15]. The connection between the physicochemical characteristics of the NP and its effectiveness in delivering tumor cells must be established [16], to solve the low tumor delivery effectiveness problem. NP tumor delivery efficiency has the roles of the physicochemical characteristics using a straightforward multiple linear regression analysis, been examined in earlier studies.[17]. The extent of these correlations was modest [18], even though some correlating variables with statistical significance are found. Due to the importance of NP tumor delivery efficiency's physicochemical properties, It is essential to address this important limitation to properly design cancer nanomedicines and increase their delivery efficiency.

3. Materials and Methods

This database included a few variables, such as the physicochemical characteristics of NP, that could affect tumor delivery. The wide range and non-normal distribution of the hydrodynamic diameter values were log-transformed. Filtering the missing data allowed the database to be reorganized to develop DL and ML models by utilizing the earlier work. [19]. During the data preprocessing stage, the database contained both categorical and numerical data to prepare the data for recognition by the DL and ML models, We used two distinct data preprocessing methods: feature scaling and one-hot encoding. The categorical variables were split into different columns, and the data was one-hot encoded for them. The artificial ordering it imposes on the variables, which could have an impact on the ML and DL models, is done using its own encoded binary string

3.1 Model Development

This study utilized a total of 9 modeling algorithms: Ensemble models, Classic models, neural networks, and support vector machines are four different classes that can be used to categorize these algorithms. The basic ML algorithms used two well-known models: k-nearest neighbors and simple linear regression. Consisting of three decision tree algorithms; the Gradient boosting model Bagged model, and random Forest, and served as an ensemble model. The linear basis kernel-based SVM models were implemented using three different iterations:least-squared, L2 Regularized, and regular SVM models.ML packages like Xgboost, Random Forest, and

kernlab were used to implement these algorithms in the R programming language for model construction. [20-21]. For each ML model, the hyperparameters were optimized using the random search technique included in the caret R package.

3.2 Performance Evaluation

To create a training set, the initial dataset was randomly divided. Through 5-fold cross-validation, external and internal training for the model has been done. Additional divisions of the training set included five subsets of equal size for the 5-fold cross-validation analysis. Out of these five subsets, one subset is retrained as validation data and the remaining 4 subsets are used for model development. Up until all subsets were used for validation once, 5 times were added to this cross-validation process. [22]. Root Mean Square Error was used to evaluate how well each model performed for the external validation and 5-fold cross-validation. The adjusted determination coefficient and mean absolute error. The following defines these evaluated metrics:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}(\sum(z - \hat{z})^2)} \quad (1)$$

$$\mathrm{MAE} = \frac{1}{n}(\sum|z - \hat{z}|) \quad (2)$$

$$R^2 = 1 - (\sum(z - \hat{z})^2 / \sum(z - \hat{z})^2) \quad (3)$$

where, using a PBM model, from the Nano-Tumor Database y represents the observed response variable value obtained

4. Results

Figure 1 shows the overall study structure. Nano-Tumor Database we have made available served as the source for all information on the properties of NPs on a physicochemical level, the nature of the tumor, and its type.

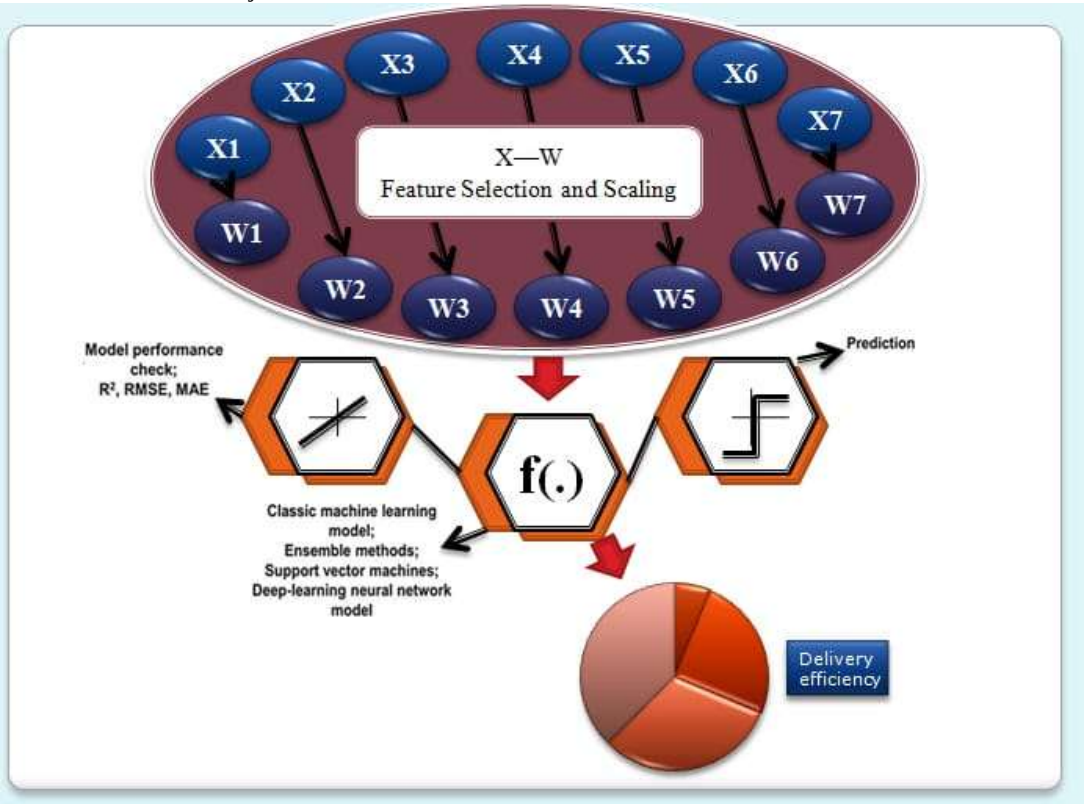

Figure 1 **The research framework's overview**

4.1 Development and Validation

compared to other categories of ML algorithms, having a lower MAE or RMSE and higher R2 values, the RF model outperformed the other chosen ML model algorithms in terms of its ability to predict each of the endpoints. The KNN model had R^2 values for all endpoints that were less than 0.1, making it the relatively weakest modeling algorithm. Each endpoint in the test set had RF R^2 and RMSE values ranging from 0.11 to 0.29 and 3.17 to 7.15, respectively, whereas the training set's values varied from 2.06 to 3.72 and 0.15 to 0.19, respectively, for each. In the test set for DE_{168}, the L2-SVM model performed slightly better than the competition. Significant variations existed between the test sets and training in the RMSE and R_2 values, which raises the possibility that the findings are not accurate. Refer to Table 3 for the DL model's performance in making predictions; With the highest R^2 values and significantly poor MAE and RMSE values in comparison to other

algorithms across all endpoints (see Figure 2), All ML method's results are outperformed. In the test set, the R^2values for DE_{max}, DE_{24}, DE_{168}, and DE_{Tlast} were 0.70, 0.46, 0.33, and 0.63, respectively. Throughout the entire training set, these values were 0.77, 0.92,0.76, and 0.77. A further indication that there were few or no overfitting issues is the similar ranges of R^2, MAE, and RMSE, between the test sets and training of the 5-fold cross-validation results in the model of DL results.

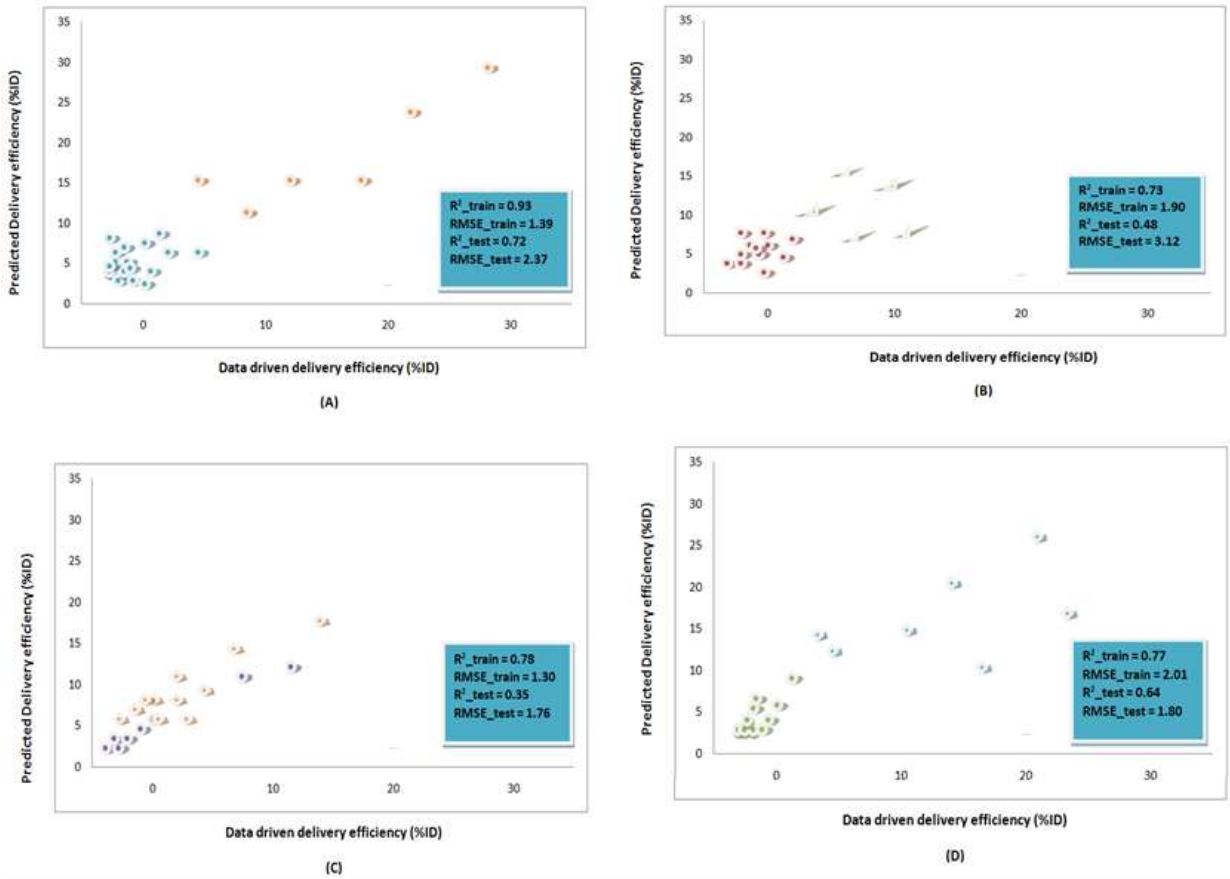

Figure 2 Relationship between values from the deep neural network model and the Nano-Tumor Database

The model of DL performance's superiority over competing models approaches and to further evaluate, Delivery efficiency was predicted using the DL model, the RF model, and the conventional simple linear regression model. The outcomes were then contrasted with those obtained using the traditional simple linear regression model. Based on R2 and RMSE, these findings demonstrated that the DL model outperformed the LM model in its ability to predict DE_{max}, DE_{24}, DE_{168}, and DE_{Tlast} (see Figure 3). Even though the training dataset's RMSE and R_2 values for the RF and DL models were similar. The DL model performed well in the dataset than the RF model. (see Figure 4). Considering test sets and training entire endpoints, these results suggest that in comparison to all other DL models, the DL model had the best predictive ability and ML models that were developed. Every developed machine learning and deep learning model's source code is available on GitHub.

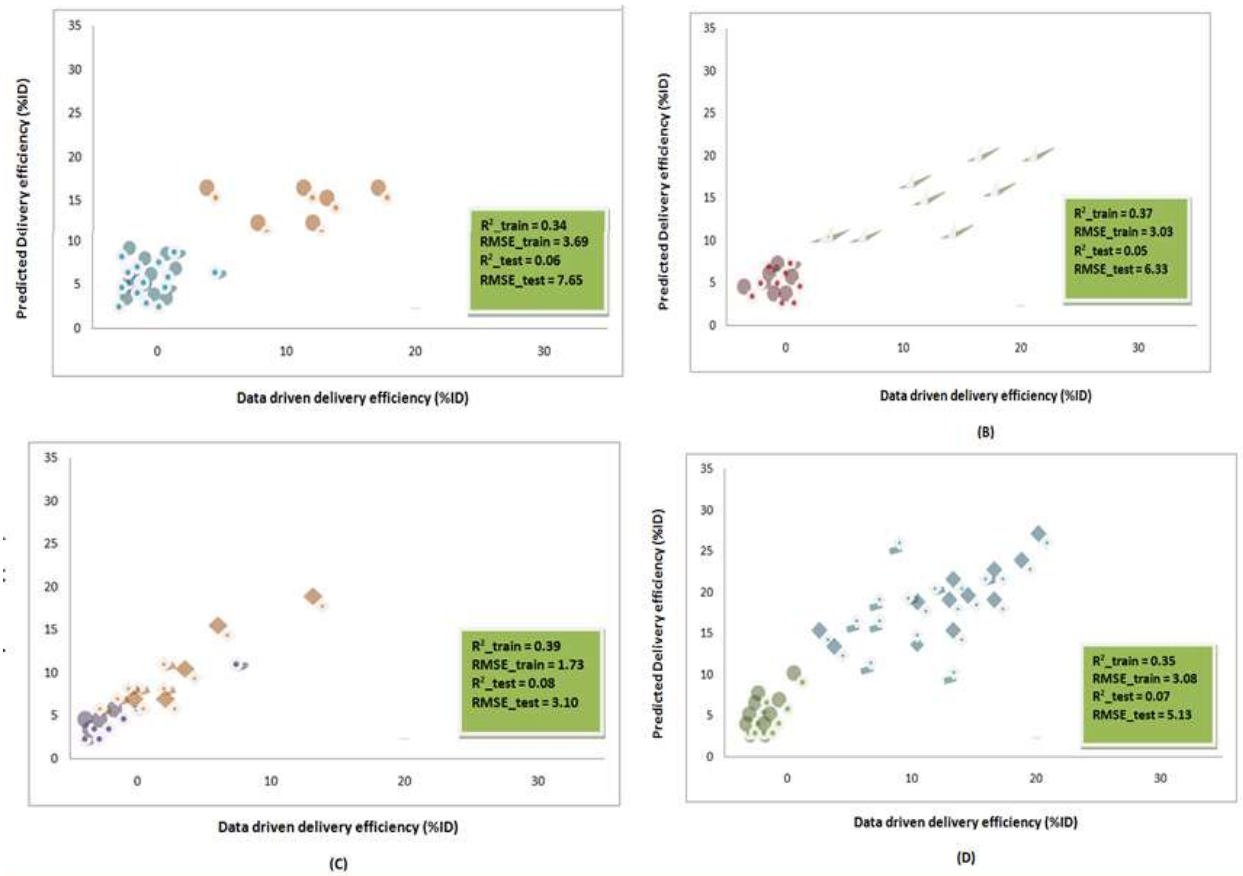

Figure 3 Relationship between values predicted by a simple linear regression model using information from the Nano-Tumor Database

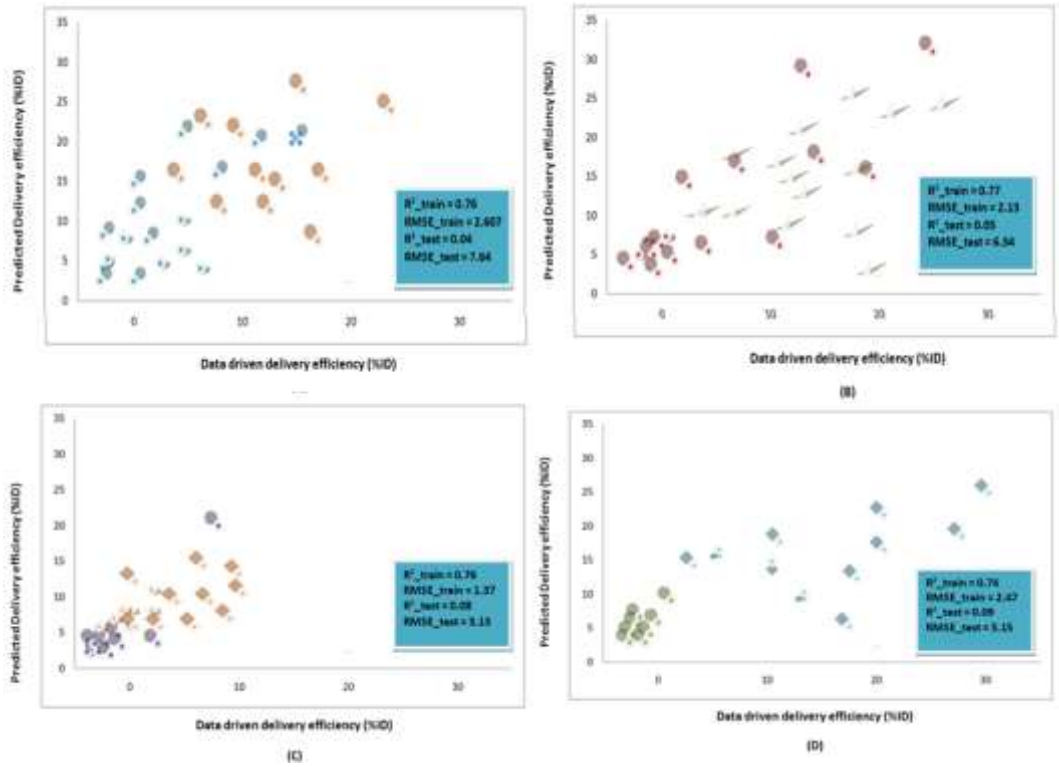

Figure 4 Correlation between Nano-Tumor Database values and values that were forecast using a random forest model

Figure 5 shows that for all endpoints, the type of cancer had a greater impact on the model of DL.MAT significantly influenced the DEmax, when it comes to the factors affecting the physicochemical characteristics of NPs, whereas ZP and Size affected the DE_{24}, DE_{168}, and DE_{Tlast}.In general, the variables of NPs were less important about the physicochemical properties than strategies of cancer therapy. The MAT and ZP were more signed up to the finished model than the rest of the NPs' physicochemical characteristics.

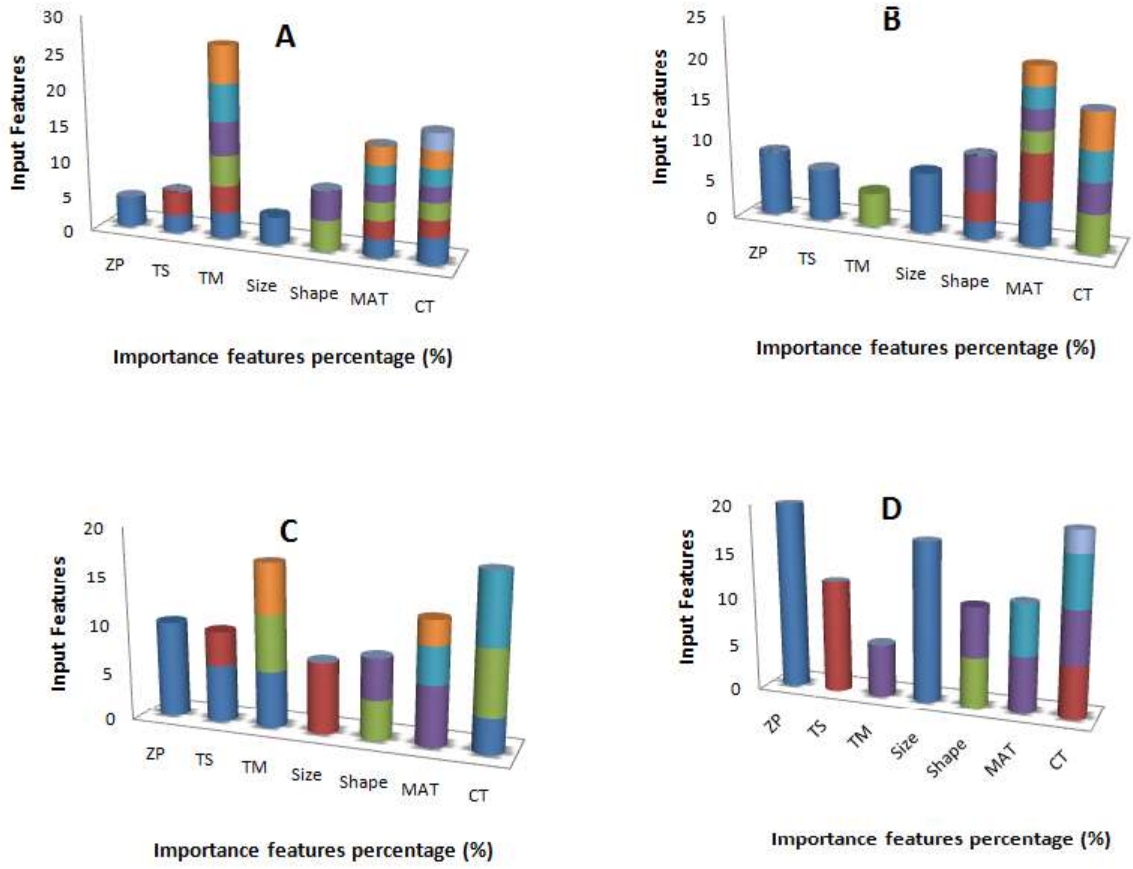

Figure 5: Percentage of each feature variable's weight in the deep learning model

4.2 Discussion

A DL neural network model based on PBM is reported in this study that cancer type, NP physicochemical characteristics, and tumor model can all be utilized to precisely forecast how well NPs will deliver their effects to mouse tumors. This model can guide the creation of new NP-based drug formulations for the treatment of cancer by using a predictive tool. preclinical trial participation must be prevented by using NP's inefficient delivery of tumors, this tool is expected to aid in the development of nanomedicine in the future. As a result, it aids in the reduction and improvement of animal studies and better-informed scientific decisions. By combining PBM modeling with AI and ML approaches, this study advances methodology in cancer nanomedicine.

Due to poor delivery effectiveness to tumors of NPs, the field of cancer nanomedicine has been hampered for a long time. We understand that a nanomedicine's efficacy and specificity are crucial elements that may compensate for less delivery efficiency. However, to reduce adverse reactions in tissues other than the target, Effectiveness will rise with improved delivery, NP treatment with similar pharmacodynamic activity may even require a lower overall dose. The relationship between tumor delivery efficiency and NP physicochemical properties was established earlier, which was conventional simple multivariate regression analysis using delivery efficiency.

All parameters were taken into account in this study, and the chosen metrics for tumor delivery efficiency were evaluated by the DL model in terms of the relative contributions of each factor under study. These findings imply that the DL model's ability to predict tumor delivery efficiency was significantly influenced by the type of cancer. The MAT and ZP were more important than the shape, type, and TS among all physicochemical properties. In the current study, the DL and ML models could be trained and evaluated thanks to

this database. The results of PBM simulations can be used in this study to guide the development of AI and ML models. Therefore, AI, PBM, and ML can be effectively integrated to support one another, according to previous studies as well as the results of the current study. The field of nanomedicine is advanced by this novel AI-PBM integrative approach. Given the numerous biomedical applications of PBM modeling, such as drug development and discovery, this conclusion has broad ramifications.

5. Conclusion

To support the development of cancer nanomedicine, the current study shows that it is feasible to combine AI/ML with PBM models. These results signify an improvement in cancer nanomedicine's methodology. The completed DL model can be used as a base for developing new cancer nanomedicines in the future and assisting researchers in making defensible decisions in-order to lessen and improve the use of animal studies to determine which NPs should enter preclinical trials. Other PBM modeling applications can be added to this framework, including the evaluation of environmental health risks, the development of small molecule drugs, and the safety of food derived from animals

.

5. References

Muñoz-Gil, G., Volpe, G., Garcia-March, M. A., Aghion, E., Argun, A., Hong, C. B., ... & Manzo, C. (2021). Objective comparison of methods to decode anomalous diffusion. Nature communications, 12(1), 1-16.

Argun, A., Volpe, G., & Bo, S. (2021). Classification, inference, and segmentation of anomalous diffusion with recurrent neural networks. Journal of Physics A: Mathematical and Theoretical, 54(29), 294003.

Park, S., Thapa, S., Kim, Y., Lomholt, M. A., & Jeon, J. H. (2021). Bayesian inference of Lévy walks via hidden Markov models. Journal of Physics A: Mathematical and Theoretical, 54(48), 484001.

Rizvi, A., Mulvey, J. T., Carpenter, B. P., Talosig, R., & Patterson, J. P. (2021). A close look at molecular self-assembly with the transmission electron microscope. Chemical Reviews, 121(22), 14232-14280.

Shiratori, K., Bishop, L. D., Ostovar, B., Baiyasi, R., Cai, Y. Y., Rossky, P. J., ... & Link, S. (2021). Machine-learned decision trees for predicting gold nanorod sizes from spectra. The Journal of Physical Chemistry C, 125(35), 19353-19361.

Kang, S., Kim, J. H., Lee, M., Yu, J. W., Kim, J., Kang, D., ... & Park, J. (2021). Real-space imaging of nanoparticle transport and interaction dynamics by graphene liquid cell TEM. Science Advances, 7(49), eabi5419.

Liu, C., Yao, L., & Chen, Q. (2022). Machine Learning Based Tracking of Single Nanoparticle Vibrations from a Projected 3D Moiré Lattice. Microscopy and Microanalysis, 28(S1), 94-95.

Latchoumi, T.P., Ezhilarasi, T.P. & Balamurugan, K. Bio-inspired weighed quantum particle swarm optimization and smooth support vector machine ensembles for identification of abnormalities in medical data. SN Appl. Sci. 1, 1137 (2019). https://doi.org/10.1007/s42452-019-1179-8

Jamali, V., &Alivisatos, A. P. (2022). Studying Diffusion of Colloidal Nanoparticles in Solution Using Liquid Phase TEM and Machine Learning. Microscopy and Microanalysis, 28(S1), 142-143.

Sung, J., Bae, Y., Park, H., Kang, S., Choi, B. K., Kim, J., & Park, J. (2022). Liquid-Phase Transmission Electron Microscopy for Reliable In Situ Imaging of Nanomaterials. Annual Review of Chemical and Biomolecular Engineering, 13, 167-191.

Cheng, D., Sha, W., Xu, Z., Huang, L., Du, Y., Tang, S., ... & Cheng, S. (2022). Computer Vision Analysis on Material Characterization Images. Advanced Intelligent Systems, 4(3), 2100158.

Gentili, A., & Volpe, G. (2021). Characterization of anomalous diffusion classical statistics powered by deep learning (CONDOR). Journal of Physics A: Mathematical and Theoretical, 54(31), 314003.

Heckert, A., Dahal, L., Tijan, R., &Darzacq, X. (2022). Recovering mixtures of fast-diffusing states from short single-particle trajectories. Elife, 11, e70169.

Park, S., Thapa, S., Kim, Y., Lomholt, M. A., & Jeon, J. H. (2021). Bayesian inference of Lévy walks via hidden Markov models. Journal of Physics A: Mathematical and Theoretical, 54(48), 484001.

Fox, Z. R., Barkai, E., &Krapf, D. (2021). Aging power spectrum of membrane protein transport and other subordinated random walks. Nature communications, 12(1), 1-9.

Li, D., Yao, Q., & Huang, Z. (2021). WaveNet-based deep neural networks for the characterization of anomalous diffusion (WADNet). Journal of Physics A: Mathematical and Theoretical, 54(40), 404003.

Reina, F., Wigg, J. M., Dmitrieva, M., Vogler, B., Lefebvre, J., Rittscher, J., &Eggeling, C. (2021). TRAIT2D: a Software for Quantitative Analysis of Single Particle Diffusion Data. F1000Research, 10.

Maizón, H. B., &Barrantes, F. J. (2022). A deep learning-based approach to model anomalous diffusion of membrane proteins: the case of the nicotinic acetylcholine receptor. Briefings in Bioinformatics, 23(1), bbab435.

Chen, Z., Geffroy, L., &Biteen, J. S. (2021). TOBIAS: Analyzing anomalous diffusion in single-molecule tracks with nonparametric Bayesian inference. Frontiers in Bioinformatics, 40.

Thapa, S., Park, S., Kim, Y., Jeon, J. H., Metzler, R., &Lomholt, M. A. (2021). Bayesian inference with scaled Brownian motion. arXiv preprint arXiv:2201.00654.

Park, S., Thapa, S., Kim, Y., Lomholt, M. A., & Jeon, J. H. (2021). Bayesian inference of Lévy walks via hidden Markov models. Journal of Physics A: Mathematical and Theoretical, 54(48), 484001.

Jamali, V., Hargus, C., Ben-Moshe, A., Aghazadeh, A., Ha, H. D., Mandadapu, K. K., &Alivisatos, A. P. (2021). Anomalous nanoparticle surface diffusion in LCTEM is revealed by deep learning-assisted analysis. Proceedings of the National Academy of Sciences, 118(10), e2017616118.