ENHANCING PRECISION IN PROTON THERAPY: UTILIZING MACHINE LEARNING FOR PREDICTING BRAGG CURVE PEAK LOCATION IN CANCER TREATMENT

Tunc ASUROGLU[1]

[1]Faculty of Medicine and Health Technology, Tampere University,
33720 Tampere, FINLAND

ABSTRACT. In proton beam therapy, the Bragg peak is the point where protons lose energy the fastest. This point is crucial for dose control, preserving healthy tissues, minimizing lateral scattering, and the success of treatment planning. However, accurately predicting the location of the Bragg peak is challenging due to the complex interactions of protons with tissues. This study proposes a machine learning (ML) approach to predict the exact location of the Bragg peak from phantom tissue proton beam therapy experiments. A dataset comprising the eight most commonly used biomaterials, which mimic human tissue in proton therapy procedures, has been curated for this study. Various ML models are benchmarked to find the most successful approach. ML model parameters are further optimized using a metaheuristic approach to achieve the highest prediction capability. In addition, feature contributions of each feature in the dataset are analyzed using an explainable artificial intelligence (XAI) technique. According to experimental results, Random Forest (RF) model that is optimized with Genetic Algorithm (GA) achieved 0.742 Correlation Coefficient (CC) value, 0.069 Mean Absolute Error (MAE) and 0.145 Root Mean Square Error (RMSE) outperforming other ML models. The proposed approach can track and predict the movement of the proton beam in real-time during treatment, enhancing treatment safety and contributing to the more effective management of the treatment process. This study is the first to predict exact Bragg curve peak locations from proton beam therapy experiments using ML approaches. The optimized ML model can provide higher precision in identifying the needed beam dosage for targeted tumor and improving treatment outcomes.

1. INTRODUCTION

Protons are positively charged and heavy, so they lose energy slowly but continuously as they scatter through matter. This scattering is at small angles [1, 2, 3]. The Bragg peak is the point where protons lose energy the fastest [4, 5]. Because tumors are large, a single-energy proton beam is not enough to treat them. Instead, proton beams with different energies are needed [4, 5, 6, 7]. It is important that the target is made of a material that is similar to tissue [8]. This is because the dose of radiation delivered to the target should be as accurate as possible. Phantoms are used to simulate the target and help optimize the dose [1, 8]. This process continues until all energies are depleted, and then they suddenly come to a halt. The dose accumulation process forms the characteristic depth-dose curve ("Bragg curve") of a broad monoenergetic proton beam. The highest dose point is called the Bragg peak. The depth of the peak, i.e., the range of protons, depends on the initial energy. Detecting the location of this peak is crucial for dose control, preserving healthy tissues, minimizing lateral scattering, and the success of treatment planning [7].

It is generally accepted that the accuracy of the results of a proton beam therapy simulation is related to the similarity between the phantom material and the tissue it is simulating [8]. The International Atomic Energy Agency (IAEA) recommends using water as the phantom material for soft tissue simulations, because it is easy to obtain and has a density similar to soft tissue [9, 10, 11]. Other biomaterials, such as those with mass densities similar to hard tissue, can also be used [12, 13, 14, 15]. The dose delivered to the target, the shape of the Bragg peak, and the results of nuclear interactions are all crucial factors to consider when evaluating the properties of a phantom material [8, 11]. Even though they are less commonly used, biomaterials are still important for simulating interactions such as backscattering, collision events, phonon production, and side scattering [1, 12].

The machine learning (ML) model, trained with information that are originated from proton beam simulations using tissue like biomaterials, can help researchers better understand the interactions of protons within tissues. ML approaches can provide higher precision in identifying the needed beam dosage for targeted tumor during treatment planning and achieving dose optimization. Accurate detection of the Bragg peak of the proton beam allows maximum focus of the treatment dose on the tumor region and enables finding methods to minimize damage to normal tissues. Additionally, the ML model can track the movement of the proton beam in real-time during treatment and quickly provide alerts in case of any deviations. This can enhance treatment safety and contributes to the more effective management of the treatment process.

In the domain of proton beam therapy, where precision and effectiveness are paramount, the importance of optimizing ML models and conducting feature impact

analysis cannot be overstated. Proton therapy, renowned for its capacity to precisely target tumors while preserving healthy tissues, stands to gain immensely from the application of ML [16, 17, 18]. The optimization of ML models becomes instrumental in refining treatment parameters, ensuring accurate predictions of pivotal factors such as the Bragg peak curve location. Through the utilization of sophisticated algorithms, clinicians can tailor ML models to accommodate diverse patient profiles, optimizing treatment plans to achieve maximum therapeutic impact. Feature impact analysis also plays a pivotal role by unveiling the influential variables that significantly impact treatment outcomes when using the proton beam therapy [19]. This insight into feature impacts enables personalized adjustments in proton beam therapy, contributing to more tailored and efficacious cancer treatments. The seamless integration of ML model optimization and feature impact analysis not only elevates the precision of proton beam therapy but also represents a substantial leap toward the implementation of individualized and optimized strategies in cancer care.

To this end, a dataset is constructed with most commonly used biomaterials, which mimics human tissue in proton therapy procedures. Afterwards various ML models are benchmarked to find out which one is better on finding the exact location of Bragg curve peaks under different energy levels and with different biomaterials. To achieve a robust model, ML model parameters are further optimized by using genetic algorithm optimization method. As a last step, feature contributions are assessed by SHAP (SHapley Additive exPlanations) technique to see which features are important when making the predictions.

Contributions of this study can be summarized as follows:

- This is the first study that aims to predict exact Bragg curve peak locations from proton beam therapy experiments using ML approaches.

- A dataset comprising the eight most commonly used biomaterials, which mimic human tissue in proton therapy procedures, has been curated for this study.

- Various ML models are benchmarked to find out which one is better on finding the exact location of Bragg curve peaks under different energy levels and with different biomaterials.

- ML model parameters are further optimized using a metaheuristic approach to achieve the highest prediction capability.

- Feature contributions of each feature in the dataset are analyzed using an explainable artificial intelligence (XAI) technique.

- By using ML approaches that are optimized by metaheuristic algorithms, this approach can track and predict the movement of the proton beam in real-time during treatment. This can enhance treatment safety and contributes to the more effective management of the treatment process.

The paper is organized in the following manner: Section 2 presents details of the proposed framework, covering the dataset, feature extraction process, machine

learning approaches, genetic optimization algorithm and feature importance assessment. Section 3 explains the evaluation metrics and highlights the experimental results. Ultimately, Section 4 concludes the paper with conclusion and discussion section.

2. Material and Methods

2.1. **Proposed machine learning framework.** Proposed ML framework leverages biomaterial features and energy levels as input and employ a ML model for establishing relationships between biomaterials and Bragg curve peak points. The framework involves a straightforward learning process, which encompasses training and testing/evaluation stages. To begin, features are extracted from particle therapy experiments, a ML model is trained using these feature vectors, where each sample corresponds to a specific peak location value. In the last step, the model predicts the peak point of a test sample. The trained model's performance is assessed using various evaluation metrics. The workflow of proposed approach is illustrated in Figure 1.
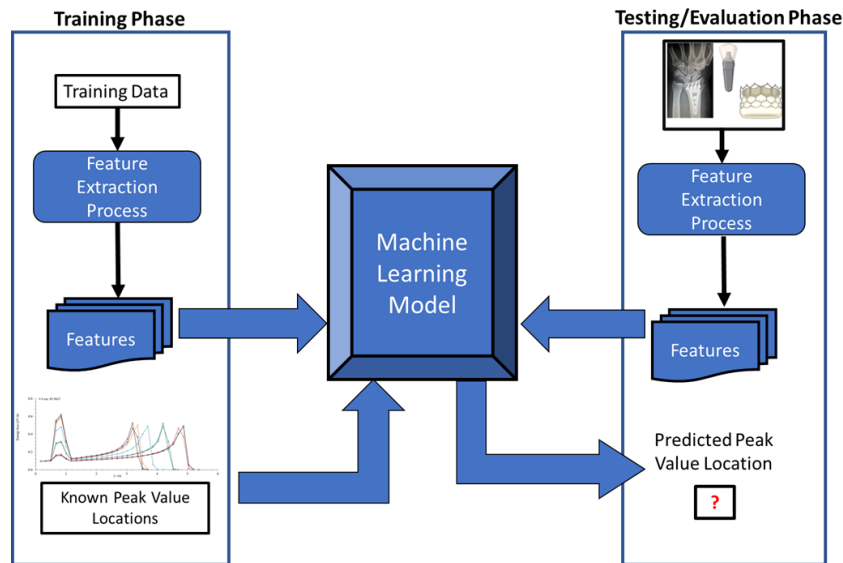

FIGURE 1. Workflow of proposed approach.

2.2. **Dataset and Feature Extraction.** In this study, the necessary data for ML models were obtained from the Monte Carlo (MC) Transport of Ions in Matter

(TRIM) simulation tool. This tool can calculate ion interactions within the target [20]. Information such as the type and energy of ions, selected phantom type and shape, incident angle of the beam, parameters to be calculated, particle count, and probability can be inputted in TRIM [20]. TRIM can calculate all kinetic events related to the energy loss processes of ions, damage inflicted on the target, scattering, ionization, voids in the crystal structure of polymeric biomaterials, phonon generation, and recoil [20]. All target atom cascades (polymeric, soft tissue, and water) in the selected phantom can be tracked and recorded in detail [21]. The recorded data aims to determine the most suitable phantom biomaterial and obtain biomaterials that are more similar to human tissue. MC TRIM and feature extraction procedure is demonstrated in Figure 2. First, experiments are conducted using MC TRIM algorithm that includes proton beam reflected on biomaterial and as a second stage biomaterial features, and Bragg curve peak locations are recorded for each experiment.
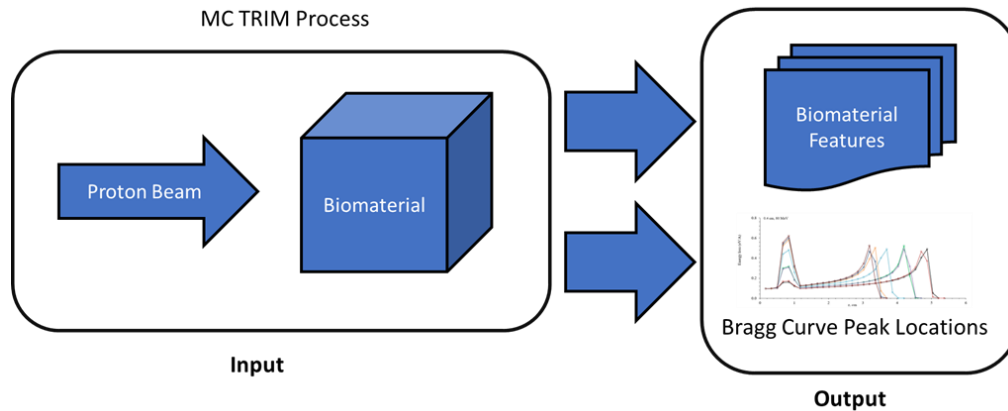

FIGURE 2. Experimental procedure to obtain Bragg curve peak locations and biomaterial features.

The experiments are conducted on 8 biomaterials: Cortical bone, Teflon, Titanium alloy, Aluminum oxide, stainless steel, Vitalium, cobalt-nickel-chromium-molybdenum, and Nital. Each biomaterial is tested with five thicknesses (0.4, 0.6, 0.8, 1, and 1.2 centimeters (cm)) and 10 different energy levels (80, 100, 120, 140, 160, 180, 200, 220, 240, and 250 Mega-electronvolt (MeV)). This combination results in a dataset of 400 samples for the machine learning models to perform predictions. Each sample in the dataset has a feature set containing information relevant to predicting the Bragg curve peak point location, which ranges between 0.23 and 1.43 cm. These features are:
- *Energy in MeV:* Energy value of the proton.

- *Biomaterial thickness (cm):* This directly affects the depth of the Bragg peak within the material.
- *Biomaterial mass density:* This property influences how much radiation the material interacts with.
- *Biomaterial atomic density:* Similar to mass density but considers the number of atoms per unit volume.
- *Biomaterial atomic composition percentage of each atom:* This captures the elemental setup of the material, which impacts its interaction with radiation. To create this feature, a feature vector is constructed for each unique atom present in the dataset. Each biomaterial sample's corresponding feature vector is filled with its respective atomic percentages. Any missing elements (not present in the specific biomaterial) are represented by zeros in the vector.

To ensure all features contribute equally to the machine learning models, the dataset undergoes normalization. This process centers the mean value of each feature to zero and scales the standard deviation to one. This standardization prevents features with larger scales from dominating the model's learning process.

2.3. **Selected Machine Learning Models for Proposed Framework.** In order to assess the effectiveness of ML models on predicting the exact peak location of the Bragg curve, various models are evaluated. Since the prediction task in this study is a regression problem due to predicting exact values instead of class labels, models that utilize regression process are selected. These models are Decision Tree (DT), Random Forest (RF), Linear Regression (LR), eXtreme Gradient Boosting (XGBoost), Support Vector Regression (SVR) and k-nearest neighbor (kNN).

The Decision Tree (DT) algorithm can manage both numerical and categorical data, seeking the feature that best divides the training set [22]. This feature is selected based on its maximum information gain. Upon evaluating the potential values of this feature, the algorithm branches into sub-trees and assigns target values. Meanwhile, it explores other features with high information gain. This iterative process continues until a clear decision is made regarding the combination of features that forms a definitive rule for predicting target values. By the end of the algorithm, all features have been assessed, and every sample has been assigned an appropriate target value [22]. Its simplicity and reliability have made it a popular ML tool across various domains [23, 24].

Random Forest (RF), a member of the decision tree family, employs an ensemble learning technique to enhance its predictive power [25]. This algorithm has gained immense popularity for its ability to effectively combat overfitting in both classification and regression tasks, while maintaining relatively low computational demands [25, 26, 27]. RF builds a multitude of decision trees by randomly sampling subsets of data, known as bootstrap samples [25]. Unlike traditional decision tree algorithms that strive to identify the optimal variable at each decision point, RF

introduces an element of randomness by considering a random subset of variables at each split. This approach is primarily implemented to mitigate the correlation among individual decision trees [25]. Such correlation can negatively impact predictions and hinder overall performance. The incorporation of randomness in RF is crucial for sound decision-making. Highly correlated variables can lead to biased predictions and suboptimal outcomes [26]. By introducing randomness, RF effectively reduces the influence of individual variables, allowing the algorithm to make more robust and unbiased decisions. The predictions from these independent decision trees are then aggregated to achieve the final outcome [26]. This ensemble approach not only addresses overfitting but also enhances the overall accuracy of the algorithm.

Linear regression (LR) stands out as one of the most fundamental and widely employed regression techniques, known for its simplicity [28]. One of its key advantages lies in the straightforward interpretability of its outcomes. In essence, linear regression fits a linear model characterized by coefficients to minimize the residual sum of squares between the observed target values in the dataset and those predicted by the linear model. Using this model, it becomes possible to make predictions for unknown target values by utilizing specified parameters along with the computed coefficients [29].

eXtreme Gradient Boosting (XGBoost) is an ensemble ML approach build on decision trees, utilizing an iterative function called gradient descent framework [30]. This iterative approach continually enhances model performance by enhancing the learning capacity of weak learners. XGBoost is versatile and applicable to both classification and regression problems, making it a valuable tool for addressing supervised learning tasks. Ensemble learning, a fundamental concept, involves creating multiple weak predictors to make predictions for a dataset and subsequently combining these individual predictions using a specific strategy to arrive at the final prediction result [31]. XGBoost represents an advancement over the traditional gradient boosting decision tree algorithm, offering improvements in terms of model building speed, prediction ability, and adjustability. In contrast to gradient boosting, XGBoost incorporates regularization within the loss function to formulate its objective function [32].

Support Vector Regression (SVR) represents a crucial facet of the broader Support Vector Machine (SVM) framework [33]. Unlike SVM classification, where multiple classes of sample points are involved, SVR is specifically tailored for situations where only one type of sample point is present. The fundamental objective of SVR differs from SVM as it does not seek to maximize the margin or separation distance between multiple types of sample points. Instead, SVR's goal is to minimize the collective deviation between the sample points and a hyperplane [34]. In the case of addressing nonlinear problems, SVR leverages a kernel function to transform the

nonlinear regression task into a higher-dimensional space. This transformation allows SVR to identify an optimal hyperplane for effectively separating the sample points in this transformed space, thereby facilitating accurate regression in cases where linear relationships may not hold [35].

The k-Nearest Neighbors (kNN) model is a supervised ML technique primarily employed for classification purposes [36]. This algorithm revolves around a flexible parameter, denoted as 'k' which represents the number of nearest neighbor's to consider when doing predictions. The kNN algorithm operates by identifying the nearest samples or neighbors within a training dataset in response to a query sample. These nearest samples are determined based on their proximity to the query sample. Once the k nearest samples are identified, the algorithm employs a majority voting rule to determine which class appears most frequently among them. The class with the highest frequency is designated as the final classification for the given query [37]. For regression problems as in this study, weighted average of the prediction value is calculated, where closer neighbors have more influence on the prediction.

2.4. **Genetic Algorithm (GA) for Machine Learning Model Parameters Estimation.** When building ML models, finding, and estimating the model parameters can be a crucial task. Improper parameter sets can lead to weak predictors. To solve this problem there are several ways, a classical approach, the grid-search is a method for parameter optimization in which a predefined range of parameter combinations is exhaustively evaluated to identify the most effective configuration for a model. Each combination in the grid is assessed for model performance using a chosen scoring metric. The set of parameters that produces the best model performance is usually chosen as the optimal configuration [38]. But this process can advance slowly and obtaining an optimal parameter set can be difficult [39]. Heuristic methods such as genetic algorithm and evolutionary algorithms can be used to find approximate solutions to computationally expensive problems more quickly and efficiently than conventional methods [40]. These methods have been used to solve numerical problems and prediction problems [41, 42]. Heuristic methods aim for feasible solutions within the problem domain. Starting from a candidate solution, they produce a new generation of solutions with modified objective values.

In this study, in order to optimize the parameters of ML models, genetic algorithm is used. In this study, the best model is selected according to the achieved results and then as a second step, GA optimization is applied to this model for further optimizing the parameter set for better performance.

GA is a type of heuristic search algorithm that mimics the process of natural selection to find optimal solutions to complex problems. It is considered as a probabilistic optimization method because GA uses randomness to explore different solutions and find the best one [43]. When employing GA, every potential solution

is represented as a chromosome within the problem's search space. Search space corresponds to the population. From a biological perspective, these chromosomes mirror the traits of an organism. The distinguishing factor is that the genes contained within different chromosomes exhibit variations. Genes are encoded using a binary encoding technique (0 or 1). Basic units of operations in GA are the chromosomes, three operators adjust the chromosomes to achieve the optimal solution. These operators are selection, crossover, and mutation operator [44].

- *Selection operator:* Throughout the course of biological evolution, all living beings must undergo adaptation to their surrounding environment, and it's only those individuals who successfully navigate and align with the demands of their habitat that emerge as superior individuals. This phenomenon is recognized as natural selection, and the selection mechanism in GA emulates this natural process. A chromosome's likelihood of elimination diminishes as its fitness score increase.

- *Crossover operator:* This operator is the most crucial one in the algorithm. It entails the exchange of gene segments from two different chromosomes, resulting in the creation of two completely new chromosomes.

- *Mutation operator:* In a biological perspective not every chromosome exchanged, some of it actually mutates to generate previously unseen new chromosomes. This probability of mutation actually helps prevent the algorithm from prematurely converging to an undesired state. It ensures that evolution is more varied and enhance the GA's capacity for local search.

In the context of GA parameter optimization, chromosome denotes the parameter values to be optimized, while the search space relates to the parameter boundaries. Within the population, chromosomes which are made of genes, represent the parameters that require optimization for the desired problem. The objective of the GA is to seek the optimal individual that meets the criteria of the fitness function. Overall working mechanism of optimization is given in Figure 3.

2.5. **Feature importance assessment using SHAP (SHapley Additive exPlanations) values.** In this study, one of the explainable artificial Intelligence (XAI) technique called SHAP is utilized to evaluate and determine the feature importance on predicting the peak point location of the Bragg curve [45]. The SHAP method is used to compute SHAP values for individual features within a ML model, enabling a better grasp of how these features impact the behavior of the model. These values are calculated for all features individually, by utilizing ML model's conditional expected value function [46].

The Shapley value is a concept within cooperative game theory that allocates the overall gains derived from collaboration among participants in a game, based on their respective incremental contributions [45]. It is shown as (1):
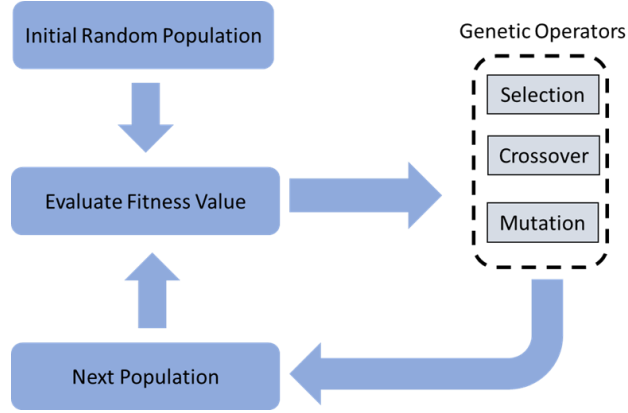
FIGURE 3. Genetic algorithm optimization process.

$$\emptyset_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(M-|z'|-1)!}{M!} [f_x(z') - f_x(z' \backslash i)] \tag{1}$$

f represents model, M corresponds to number of x', whereas x' represents a condensed input that corresponds to the initial input via a mapping function $x = h_x(x')$. Here h_x assigns 1 or 0 to the initial inputs, 0 means that input is not considered for the model whereas 1 means otherwise. $|z'|$ represents the count of non-zero elements within z', and $z' \subseteq x'$ denotes all vectors of z' where the non-zero elements form a subset of the non-zero elements in x' [45]. This process calculates a significance value for each feature, and it corresponds to the effect on prediction of the model. The importance of the ith feature is calculated by comparing the predictions of two models: one trained with all features, and the other trained without the ith feature. As the impact of excluding a feature is contingent on the presence of other features in the model, the earlier differences are calculated for every potential subset denoted as $z' \backslash i$. Thus, the Shapley value can be characterized as a distinctive method for attributing features that computes a weighted average of all potential differences [46].

ML SHAP value is created in a similar manner with the traditional SHAP value, utilizing the conditional expectations to designate mapped inputs [47]. In ML SHAP value calculation, conditional expectation $(E[f(z)|z_S)$ is used instead of $f_x(z')$ to acquire the SHAP values. Here, S represents the collection of indices contained within z. For every sample in the dataset, SHAP values are calculated and therefore contribution of each feature on model prediction can be analyzed [48]. The conventional method for calculating feature importance is to average the absolute

values of the SHAP values that are calculated for all instances. It is shown as (2) (N represents the number of samples in the dataset):

$$S_0 = \frac{1}{N}\sum_{i=1}^{N}|\emptyset_i| \tag{2}$$

3. RESULTS

3.1. **Experimental Setup.** The dataset is divided into two parts, one part for training and validation, the other part is for testing. The division ratio is 70% for training and validation, 30% for testing. In GA optimization phase, best model parameters are determined using a tenfold cross-validation (CV) approach. In this approach, the dataset is divided into ten equal parts, with one part reserved for validation while the remaining nine parts serve as the training set. The CV process concludes after each part has been used as the validation set. The CV approach is used to determine the best trained ML model based on available training data. After that the trained model is evaluated on a test set that is not seen by the model before and experimental results are reported afterwards. This train-validate-test approach measures generalization ability of ML models on an unseen test dataset.

Regression performance of ML models are evaluated in terms of Correlation Coefficient (CC), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics. These metrics assess how well the regression models behave when predicting exact Bragg curve peak point location values. To attain a high level of performance, the model should demonstrate low error rates, alongside a high correlation value.

CC is a metric that ranges from -1 to 1. Positive correlation is depicted as +1 whereas negative correlation represented by -1. CC metric is consisting of various variables, n refers to dataset sample size, a_i and p_i variables refer to actual and predicted values, respectively. $\bar{a}$ and $\bar{p}$ are calculated mean values of actual and predicted values. The correlation between estimated and actual values is determined using the CC value. It is shown as (3):

$$CC = \frac{S_{PA}}{\sqrt{S_P S_A}} \tag{3}$$

$$S_{PA} = \frac{\sum_i(p_i - \bar{p})(a_i - \bar{a})}{n-1}$$
$$S_P = \frac{\sum_i(p_i - \bar{p})^2}{n-1}, \text{ and } S_A = \frac{\sum_i(a_i - \bar{a})^2}{n-1}.$$

Another metric, MAE quantifies the disparity between two continuous variables. It is given as (4):

$$MAE = \frac{|p_1 - a_1| + \cdots + |p_n - a_n|}{n} \tag{4}$$

RMSE, a quadratic metric, serves as a reliable measure of the error magnitude in machine learning models. It effectively quantifies the discrepancy between the model's predicted values and the actual observed values (5):

$$RMSE = \sqrt{\frac{(|p_1 - a_1|)^2 + \cdots + |p_n - a_n|^2}{n}} \tag{5}$$

3.2. **Experimental Results.** Various experiments are performed on existing ML models. These models are Decision Tree (DT), Random Forest (RF), Linear Regression (LR), eXtreme Gradient Boosting (XGBoost), Support Vector Regression (SVR) and k-nearest neighbor (kNN). The parameter setup for selected ML models is given in Table 1.

The experimental results for regression models aimed at predicting the peak location values are summarized in Table 2. The RF demonstrated superior performance, achieving correlation coefficient (CC), mean absolute error (MAE), and root mean square error (RMSE) values of 0.712, 0.073, and 0.151, respectively. As can be seen from Table 2, RF model outperforms other regression models across all evaluation metrics. The DT model ranked second in terms of CC and RMSE. XGBoost also ranked second in terms of MAE. In contrast, SVR performed worst in terms of CC, MAE and RMSE values. From these results, we can infer that RF model can accurately map the relationships between biomaterial features and energy levels when predicting Bragg curve peak point locations. Another conclusion that can be drawn from these results is that overall, the models that utilize decision tree architectures performed well when making predictions.

In order to further increase the performance of the best prediction model, metaheuristic GA optimization is applied. In this experiment, GA is used to optimize the hyperparameters of RF. Optimized hyper parameters are number of trees, minimum samples per split, and minimum samples in leaf. Results of this experiment can be seen in Table 3. Different GA population (P) sizes are benchmarked to determine the best optimization approach to tune the hyperparameters of RF. According to Table 3. RF algorithm that is tuned with the population size of 50 achieved highest CC (0.742), lowest MAE (0.069) and RMSE (0.145) values. RF with population size selected as 100 performed worst in terms of CC (0.627), MAE (0.081) and RMSE (0.17). As can be seen from Table 3. GA improved the overall capacity of RF model when predicting the exact values of Bragg curve peak point locations.

TABLE 1. Parameter setup for machine learning models.

Model	Parameters
RF	Number of trees = 100
	Minimum samples per split = 2
	Minimum samples in leaf = 1
	Split criteria = MSE
LR	-
XGBoost	Objective = MSE
	Number of trees=100
	Learning rate = 0.3
	Maximum depth = 6
	Lambda = 1
DT	Split criteria = MSE
	Minimum samples per split = 2
	Minimum samples in leaf = 1
SVR	Kernel = rbf
	C = 1
	Epsilon = 0.1
kNN	k = 3

TABLE 2. Experimental results for regression models aimed at predicting peak value.

Model	CC	MAE	RMSE
RF	**0.712**	**0.073**	**0.151**
LR	0.554	0.103	0.179
XGBoost	0.618	0.078	0.177
DT	0.682	0.081	0.167
SVR	0.517	0.12	0.186
kNN	0.567	0.09	0.178

In order to highlight the positive effect of GA algorithm on RF model, Figure 4 and Figure 5 are given. In Figure 4, it can be seen that in terms CC values, "RF with GA" outperformed "RF without GA". Same situation is also can be said in terms of error metrics. In Figure 5, "RF with GA" had low MAE and RMSE values compared with "RF without GA".

TABLE 3. Performance comparison for different population sizes using genetic algorithm.

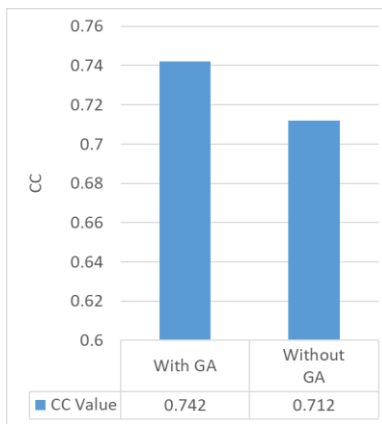| Model	CC	MAE	RMSE
RF	0.712	0.073	0.151
RF + GA (P. size=10)	0.666	0.077	0.162
RF + GA (P. size=50)	**0.742**	**0.069**	**0.145**
RF + GA (P. size=100)	0.627	0.081	0.17

FIGURE 4. Comparison of RF model with applying GA and without GA in terms of CC.
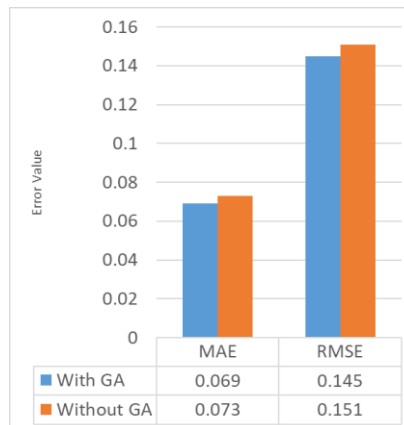

FIGURE 5. Comparison of RF model with applying GA and without GA in terms of MAE and RMSE.

As a final analysis, impact of biomaterial features on Bragg curve peak prediction is evaluated. This evaluation is done by SHAP analysis of "RF + GA (P. size=50)" ML model. In order to demonstrate this analysis, SHAP values of all samples are calculated based on their features.

Summary plots play a crucial role in SHAP analysis, providing not only the hierarchy of input variable importance but also illustrating their correlation with the target variable. One type of summary plot, the beeswarm plot, is illustrated in Figure 6. The beeswarm plot consolidates the SHAP values of all samples, enabling the simultaneous visualization of these values. The summary plot displays input variables and their importance in descending order on the y-axis, while the x-axis represents specific SHAP values. Dot color signifies magnitude (blue for small, red for large), with each dot representing a dataset sample. The horizontal x-axis illustrates the variation in SHAP values for each variable, ranging from blue to red, indicating the shift in input variable magnitude and its impact on prediction. It can be clearly observed that energy, biomaterial mass density, and biomaterial thickness features make a substantial contribution to Bragg curve peak point location prediction. According to Figure 6, lower energy levels result in higher Bragg curve peak point locations, while increased biomaterial mass densities and thickness also lead to higher Bragg curve peak point locations.

It is interesting to note that the model identifies lower energy levels as contributing to higher Bragg curve peak point locations. This seems counterintuitive based on established principles in proton therapy, where lower energy typically leads to shallower peaks [2]. This can be caused by the model might not perfectly capture the complex relationship between energy and Bragg peak location, especially at the lower energy levels. Denser materials require higher energies for protons to reach a specific depth. The model reflects this by predicting a deeper Bragg peak with increasing mass density. This helps tailor the proton beam energy to ensure the peak coincides with the tumor location within the patient's body, composed of tissues with varying densities [3]. Similar to mass density, a thicker material necessitates higher energy protons to achieve a deeper Bragg peak placement. The model's prediction aligns with this principle, allowing for treatment planning that considers the target depth within the patient's specific anatomy. By understanding these feature relationships, the model's predictions can be leveraged to optimize treatment delivery in proton therapy. This translates to more precise targeting of tumors while minimizing radiation exposure to healthy tissues.

4. DISCUSSION AND CONCLUSION

This study aims to utilize ML approaches to accurately predict Bragg curve peak locations in proton beam therapy. The development of a benchmark dataset

encompassing biomaterials that are identical to the human tissue in cancer proton therapies establishes a solid footing for evaluating various ML models. It delves into the effectiveness of different models across varying energy levels and biomaterial characteristics, further employing GA optimization to refine model, thereby bolstering predictive prowess. According to experimental results, the RF model demonstrated superior performance, achieving correlation coefficient (CC), mean absolute error (MAE), and root mean square error (RMSE) values of 0.712, 0.073, and 0.151, respectively. RF model outperforms other regression models across all evaluation metrics. It can be inferred from this result that RF model can map relationships between biomaterial characteristics and energy levels using multiple random tree architectures. Further optimizing RF model parameters using GA is proven to increase the performance of predicting exact value of Bragg curve peak location. RF model with optimized parameters outperformed the model without optimization in terms of 0.742 CC, 0.069 MAE and 0.145 RMSE, respectively.

The study's other main contribution is using SHAP method to dissect the intricate web of feature contributions. This analysis unveils the importance of each feature in predicting Bragg curve peak locations, providing invaluable insights into the factors that profoundly affect treatment outcomes. It can be clearly seen from SHAP analysis that energy, biomaterial mass density, and biomaterial thickness make a substantial contribution to Bragg curve peak point location prediction. Lower energy levels higher Bragg curve peak point locations, while increased biomaterial mass densities and thickness also lead to higher Bragg curve peak point locations. This insight into feature impacts enables personalized adjustments in proton beam therapy, contributing to more tailored and efficacious cancer treatments. The seamless integration of ML model optimization and feature impact analysis not only elevates the precision of proton beam therapy but also represents a substantial leap toward the implementation of individualized and optimized strategies in cancer care.

Through the utilization of these approaches, clinicians can tailor ML models to accommodate diverse patient profiles, optimizing treatment plans to achieve maximum therapeutic impact.

This study has certain limitations, with the primary concern being the computational load. Despite the effectiveness of RF models in various tasks, their computational complexity becomes significant, especially when dealing with a large number of samples and features. The integration of GA optimization can further exacerbate computation time in the quest for the optimal parameter combination. To alleviate this burden, one approach is to employ feature selection and Principal Component Analysis (PCA) methods, which can effectively reduce the dimensionality of features, thereby expediting the training phase.
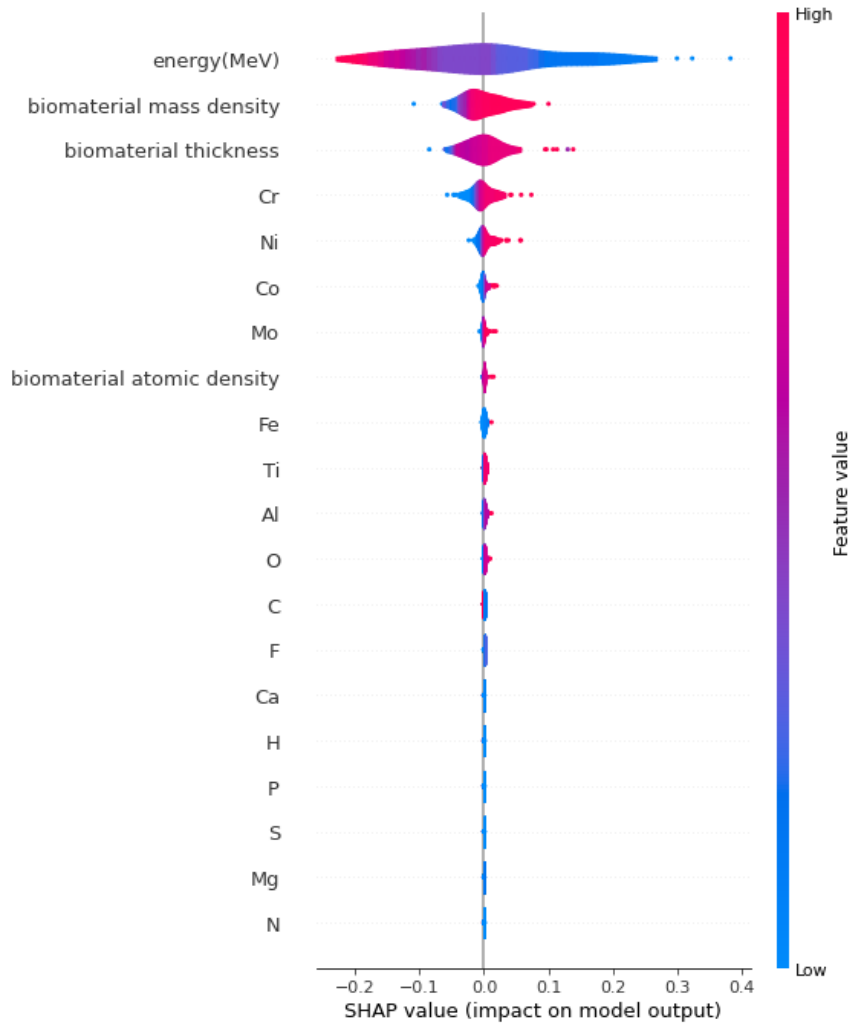
FIGURE 6. Feature importance analysis by SHAP.

An alternative approach involves harnessing Graphics Processor Units (GPUs) within a sophisticated centralized computer system to accelerate the training of the RF algorithm. Additionally, exploring parallel processing techniques can be a viable strategy to address and mitigate this challenge. Another constraint arises from the limited number of samples; given the small sample size, it is impractical to benchmark and study deep learning approaches within the proposed framework using this dataset. Deep learning methods typically require substantial datasets for

effective training. To address these limitations, a potential solution is to conduct additional experiments involving diverse biomaterials and energy levels to augment the sample size. It can increase the performance of proposed framework because the model can train with more varied data. Potential biases within the dataset might not be fully captured by the current analysis. Expanding the dataset with a broader range of biomaterials and energy levels could mitigate bias and improve the model's generalizability to real-world clinical scenarios. While the study demonstrates the effectiveness of the model in predicting Bragg peak location, further work is necessary to translate these findings into clinically applicable tools. Evaluating the model's performance on real-world patient data with complex anatomies would be crucial for establishing its clinical relevance. Also, developing a seamless integration pathway for the model's predictions into existing treatment planning software used by clinicians would enhance its practical usability.

For the future direction of this study, other potential problems, such as dose calculation and treatment planning optimization, which are crucial in particle therapy, will be investigated. The applicability of ML methods to these problems will be explored thoroughly. Various XAI techniques can be explored to understand how these advanced models arrive at their predictions, fostering acknowledgement from medical professionals. Multimodal approaches that consider patient, biomaterial, imaging data can also be integrated for better prediction capability. Especially, multimodal transformers can be adopted to learn joint representations across different data modalities, potentially leading to more comprehensive and informative features for improved prediction capability. Another future task could be increasing the sample size of the curated dataset by conducting more experiments with different biomaterials and energy levels. This approach could open up the possibility of including deep learning methods. Lastly, different hyperparameter optimization algorithms can also be considered to determine which one is better suited for this problem.

Declaration of Competing Interests The author declares no known competing interests.

References

[1] Ekinci, F., Bölükdemir, M. H., The effect of the second peak formed in biomaterials used in a slab head phantom on the proton Bragg peak, *J. Polytech.,* 23 (1) 2020, 129-136, http://doi.org/10.2339/politeknik.523001.

[2] Ekinci, F., Bostancı, G. E., Dağlı, Ö., Güzel, M. S., Analysis of Bragg curve parameters and lateral straggle for proton and carbon beams, *Commun. Fac. Sci.Univ. Ank. Series*

A2-A3: Phys. Sci. and Eng., 63 (1) (2021), 32-41, https://doi.org/10.33769/ aupse.864475.

[3] Ekinci, F., Bostanci, E., Güzel, M. S., Dagli, O., Effect of different embolization materials on proton beam stereotactic radiosurgery arteriovenous malformation dose distributions using the Monte Carlo simulation code, *J. Radiat. Res. App. Sci.*, 15 (3) 2022, 191-197, https://doi.org/10.1016/j.jrras.2022.05.011.

[4] Gottschalk, B., Proton Therapy Physics, Taylor & Francis Inc., USA, 2012, https://doi.org/10.1201/b22053.

[5] Ekinci, F., Bostanci, E., Güzel, M. S., Dağli, Ö., Analysing the effect of a cranium thickness on a Bragg peak range in the proton therapy: a TRIM and GEANT4 based study, *St. Petersbg. State Polytech. Univ. J.: Phys. Math.*, 15 (2) (2022) 64-78, https://doi.org/0.18721/JPM.15207.

[6] Carlsson, A. K., Andrea, P. and Brahme, A., Monte Carlo and analytical calculation of computerized treatment plan optimization, *Phys. Med. Biol.,* 42 (1997), 1033-1053, https://doi.org/10.1088/0031-9155/42/6/004.

[7] Hall, E. J., Kellerer, A. M., Rossi, H. H., Lam, Y-M.P., The relative biological effectiveness of 160 MeV protons-II, *Int. Rad. Onc. Biol. Phys.*, 4 (1978), 1009-1013, https://doi.org/10.1016/0360-3016(78)90013-5.

[8] Lourenço, A., Wellock, N., Thomas, R., Homer, M., Bouchard, H., Kanai, T., MacDougall, N., Royle, G., Palmans, H., Theoretical and experimental characterization of novel water-equivalent plastics in clinical high-energy carbon-ion beams, *Physics in Medicine and Biology*, 61 (21) (2016), 7623-7638, https://doi.org/10.1088/0031-9155/61/21/7623.

[9] Arib, M., Medjadj, T., Boudouma, Y., Study of the influence of phantom material and size on the calibration of ionization chambers in terms of absorbed dose to water, *J. Appl. Clin. Med. Phys.*, 7 (2006), 55-64, https://doi.org/10.1120/jacmp.v7i3.2264.

[10] Samson, D. O., Jafri, M. Z. M., Shukri, A., Hashim, R., Sulaiman, O., Aziz, M. Z. A., Yusof, M. F. M., Measurement of radiation attenuation parameters of modified defatted soy flour-soy protein isolate-based mangrove wood particleboards to be used for CT phantom production, *Radiat. Environ. Biophys.*, 59 (2020), 483-501, https://doi.org/ 10.1007/s00411-020-00844-z.

[11] Kanematsu, N., Koba, Y., Ogata, R., Evaluation of plastic materials for range shifting range compensation and solid phantom dosimetry in carbon-ion radiotherapy, *Med. Phys.*, 40 (2013), 041724, https://doi.org/10.1118/1.4795338.

[12] Senirkentli, G. B., Ekinci, F., Bostanci, E., Güzel, M. S., Dağli, Ö., Karim, A. M., Mishra, A., Therapy for mandibula plate phantom, *Healthcare*, 9 (167) (2021), https://doi.org/10.3390/ healthcare9020167.

[13] Ekinci, F., Investigation of tissue equivalence of phantom biomaterials in 4He heavy ion therapy, *Radiat. Eff. Defects Solids*, 178 (3-4) (2023), 500-509,

https://doi.org/10.1080/10420150.2022.2153251.

[14] Ekinci, F., Asuroglu, T., Acici, K., Monte Carlo simulation of TRIM algorithm in ceramic biomaterial in proton therapy, *Materials*, 16 (13) (2023), 4833, https://doi.org/10.3390/ma16134833.

[15] Ekinci, F., Bostanci, E., Güzel, M. S., Dagli, Ö., A Monte Carlo study for soft tissue equivalency of potential polymeric biomaterials used in carbon ion radiation therapy, *Nucl. Technol.*, 209 (8) (2023), 1-11, https://doi.org/10.1080/ 00295450.2023.2188144.

[16] Borderias-Villarroel, E., et al., Machine learning-based automatic proton therapy planning: Impact of post-processing and dose-mimicking in plan robustness, *Med. Phys.*, 50 (2023), 4480-4490, https://doi.org/10.1002/mp.16408.

[17] Lerendegui-Marco, J., et al., Towards machine learning aided real-time range imaging in proton therapy, *Sci. Rep.,* 12 (2022), 2735, https://doi.org/10.1038/s41598-022-06126-6.

[18] Chang, C. W., Validation of a deep learning-based material estimation model for Monte Carlo dose calculation in proton therapy, *Phys. Med. Biol.*, 67 (21) (2022), 215004, https://doi.org/10.1088/1361-6560/ac9663.

[19] Chen, Y., et al., Understanding machine learning classifier decisions in automated radiotherapy quality assurance, *Phys. Med. Biol.*, 67 (2022), 025001, https://doi.org/10.1088/1361-6560/ac3e0e.

[20] Foster, D. G., Artur, E. D., Avarege Neutronic Properties of "Prompt" Fission Products, *Los Alamos National Laboraty Report*, LA--9168-MS (1982).

[21] Ziegler, J. F., SRIM: The stopping and range of ion in matter (2013). Available at: http://www.srim.org/. [Accessed November 2023].

[22] Bhat, P., Malaganve, P., Effect of J48 and LMT algorithms to classify movies in the web-a comparative approach, Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems, Springer Singapore, 2021, https://doi.org/10.1007/978-981-33-4543-0_58.

[23] Ilyas, H., et al., Chronic kidney disease diagnosis using decision tree algorithms, *BMC Nephrol.,* 22 (1) 2021, 273, https://doi.org/10.1186/s12882-021-02474-z.

[24] Rahmayanti, N., Pradani, H., Pahlawan, M., Vinarti, R., Comparison of machine learning algorithms to classify fetal health using cardiotocogram data, *Procedia Comp. Sci.*, 197 (2022), 162-171, https://doi.org/10.1016/j.procs.2021.12.130.

[25] Simsekler, M. C. E., Alhashmi, N. H., Azar, E., King, N., Luqman, R., Al Mulla, A., Exploring drivers of patient satisfaction using a random forest algorithm, *BMC Med. Inform. Decis. Mak.,* 21 (1) (2021), 157, https://doi.org/10.1186/s12911-021-01519-5.

[26] Açıcı, K., Erdaş, Ç. B., Aşuroğlu, T., Toprak, M. K., Erdem, H., Oğul, H., A random forest method to detect Parkinson's Disease via gait analysis, Communications in Computer and Information Science, Springer, Switzerland, 2017, https://doi.org/10.1007/978-3-319-65172-9_5.

[27] Aşuroğlu, T., Oğul, H., A deep learning approach for sepsis monitoring via severity score estimation, *Comput. Methods Programs in Biomed.* 198 (2021), 105816, https://doi.org/10.1016/j.cmpb.2020.105816.

[28] Oyeleye, M., Chen, T., Titarenko, S., Antoniou, G., A predictive analysis of heart rates using machine learning techniques, *Int. J. Environ. Res. Public Health,* 19 (2022), 2417, https://doi.org/10.3390/ijerph19042417.

[29] Huang, L., Song, T., Jiang, T., Linear regression combined KNN algorithm to identify latent defects for imbalance data of ICs, *Microelectron. J.*, 131 (2023), 105641, https://doi.org/10.1016/j.mejo.2022.105641.

[30] Wu, J., et al., Prediction and screening model for products based on fusion regression and XGBoost classification, *Comput. Intell. Neurosc.*, (2022), https://doi.org/10.1155/2022/4987639.

[31] Shin, H., XGBoost regression of the most significant photoplethysmogram features for assessing vascular aging, *IEEE J. Biomed. Health Inform.*, 26 (7) (2022), 3354-3361, https://doi.org/10.1109/JBHI.2022.3151091.

[32] Manoharan, A., Begam, K. M., Aparow, V. R., Sooriamoorthy, D., Artificial neural networks, gradient boosting and support vector machines for electric vehicle battery state estimation: A review, *J. Energy Storage*, 55 (A) (2022), 105384, https://doi.org/10.1016/j.est.2022.105384.

[33] Quan, Q., et al., Research on water temperature prediction based on improved support vector regression, *Neural Comput. & Applic.,* 34 (2022), 8501-8510, https://doi.org/10.1007/s00521-020-04836-4.

[34] Nilashi, M., Abumalloh, R. A., Minaei-Bidgoli, B., Samad,S., Ismail, M. Y., Alhargan, A., Zogaan, W. A., Predicting Parkinson's Disease progression: Evaluation of ensemble methods in machine learning, *J. Health. Eng.*, (2022), https://doi.org/10.1155/2022/2793361.

[35] Sharin, S. N., Radzali, M. K., Sani, M. S. A., A network analysis and support vector regression approaches for visualising and predicting the COVID-19 outbreak in Malaysia, *Health. Anal.*, 2 (2022), 100080, https://doi.org/10.1016/j.health.2022.100080.

[36] Uddin, S., et al., Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction, *Sci. Rep.*, 12 (2022), 6256, https://doi.org/10.1038/s41598-022-10358-x.

[37] Lin, G., Lin, A., Gu, D., Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient, *Infor. Sci.*, 608 (2022), 517-531, https://doi.org/10.1016/j.ins.2022.06.090.

[38] Fayed, H. A., Atiya A. F., Speed up grid-search for parameter selection of support vector machines, *Appl. Soft Comput.*, 80 (2019), 202-210, https://doi.org/ 10.1016 /j.asoc.2019. 03.037.

[39] Sun, Y., et al., An improved grid search algorithm to optimize SVR for prediction, *Soft Comput.,* 25 (2021), 5633-5644, https://doi.org/10.1007/s00500-020-05560-w.

[40] Hamdia, K. M., Zhuang, X., Rabczuk, T., An efficient optimization approach for designing machine learning models based on genetic algorithm, *Neural Comput. and Applic.*, 33 (2021), 1923-1933, https://doi.org/10.1007/s00521-020-05035-x.

[41] Da, L., Sun, K., Random forest solar power forecast based on classification optimization, *Energy*, 187 (2019), 115940, https://doi.org/10.1016/ j.energy.2019.115940.

[42] Chui, K. T., Gupta, B. B., Vasant, P., A genetic algorithm optimized RNN-LSTM model for remaining useful life prediction of turbofan engine, *Electronics*, 10 (3) (2021), 285, https://doi.org/10.3390/electronics10030285.

[43] Beyaz, S., Açıcı, K., Sümer, E., Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches, *Jt. Dis. Relat. Surg.*, 31 (2) (2020), 175-183, https://doi.org/10.5606/ehc.2020.72163.

[44] Zhou, J., et al., Employing a genetic algorithm and grey wolf optimizer for optimizing RF models to evaluate soil liquefaction potential, *Artif. Intell. Rev.,* 55 (2022), 5673-5705, https://doi.org/10.1007/s10462-022-10140-5.

[45] Lee, Y. G., et al., SHAP value-based feature importance analysis for short-term load forecasting, *J. Electr. Eng. Technol.,* 18 (2023), 579-588, https://doi.org/10.1007/s42835-022-01161-9.

[46] Gramegna, A., Giudici, P., Why to buy insurance? An explainable artificial intelligence approach, *Risks,* 8 (4) (2020), 137, https://doi.org/10.3390/risks8040137.

[47] Kim, Y., Kim, Y., Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models, *Sustain. Cities Soc.*, 79 (2022), 103677, https://doi.org/10.1016/j.scs.2022.103677.

[48] Alenezi, R., Ludwig, S. A., Explainability of cybersecurity threats data using SHAP, *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, (2021), 01-10, https://doi.org/10.1109/SSCI50451.2021.9659888.