A physically constrained Energy Spectrum Superposition Method-Machine Learning coupling algorithm for obtaining the neutron spectrum of BNCT rapidly

Authors: Yang Junkai, Hao-PengDeng, Zhi-Meng Hu, Fang-Cong Zhang, Li Kaiguo, MinPeng, Deng Jiexiao, Ping-Quan Wang, Hui Zhang, Bowen Zhou, ChungmingPaulChu, Lin Xiao, GiuseppeGorini

Date: 2025-09-26T16:36:59+00:00

Abstract

Accurate and rapid acquisition of neutron energy spectra is one of the major challenges in image-guided radiation therapy for boron neutron capture therapy (IGRT-BNCT). This study integrates machine learning (ML), including deep neural network (DNN) and random forest (RF) algorithms, into the energy spectrum superposition method (ESSM) to predict unknown spectra not present in the database, referred to as ESSM-ML. The new algorithm can further improve the speed of spectrum acquisition while maintaining high precision. Both DNN and RF demonstrate high R^2 values and extremely low root mean square error (RMSE) for neutron energy spectrum predictions from two physical processes. The deviation between predicted and true values is small across three energy regions: thermal neutrons, epithermal neutrons, and fast neutrons. For the total neutron energy spectrum at the emission window of the beam shaping assembly (BSA), the ratio of the total neutron fluence rate obtained by ESSM-ML to that obtained by conventional simulation-based methods is 95.3%. Furthermore, the time required for ESSM-ML to obtain therapeutic spectra is only 69 seconds, representing a 4500-fold improvement in computational efficiency. The average prediction time of the ML module for a single spectrum is only 0.0052 seconds. ESSM-ML provides a theoretical and algorithmic foundation for implementing IGRT-BNCT.

Full Text

Preamble

A Physically Constrained Energy Spectrum Superposition Method-Machine Learning Coupling Algorithm for Rapid Neutron Spectrum Acquisition in BNCT

Jun-Kai Yang,[1],[2] Hao-Peng Deng,[1],[2] Zhi-Meng Hu,[3] Fang-Cong Zhang,[1] Li-Kai Guo,[1] Min Peng,[1] Deng-Jie Xiao,[4] Ping-Quan Wang,[5] Hui Zhang,[5] Bo-Wen Zhou,[6] Chungming Paul Chu,[1],† Lin Xiao,[7] and Giuseppe Gorini[8]

[1]College of Engineering and Applied Sciences, Nanjing University, Nanjing 210093, China
[2]These authors contributed equally to this work and should be considered co-first authors
[3]Department of Nuclear Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210093, China
[4]Shenzhen Polytechnic University, The Tech X Academy, Shenzhen 518000, China
[5]National Institute of Metrology, Beijing 100029, China
[6]Suzhou Laboratory, Suzhou 215000, China
[7]School of Physics, Nanjing University, Nanjing 210093, China
[8]Department of Physics, University of Milan-Bicocca, Milan 20126, Italy

Accurately and rapidly obtaining the neutron energy spectrum represents one of the most significant challenges for Image-Guided Radiotherapy of BNCT (IGRT-BNCT). In this paper, machine learning (ML) techniques, including deep neural networks (DNN) and random forest (RF) algorithms, are integrated into the energy spectrum superposition method (ESSM) to predict unknown energy spectra not present in the database, resulting in the ESSM-ML framework. This novel algorithm substantially improves spectrum acquisition speed while maintaining high accuracy. Predictions of neutron energy spectra for two physical processes using both DNN and RF demonstrate high R^2 values and extremely low RMSE. Across the three energy regions—thermal, epithermal, and fast neutrons—the deviations between predicted and true values remain minimal. For the total neutron energy spectrum at the beam shaping assembly (BSA) emission window, the ratio of total neutron fluence rate obtained by ESSM-ML to that from traditional simulation-based methods reaches 95.3%. Moreover, the time required to obtain treatment energy spectra using ESSM-ML is only 69 seconds, representing a 4,500-fold improvement in computational efficiency. The average prediction time for a single energy spectrum by the ML modules is merely 0.0052 seconds. ESSM-ML provides the theoretical and algorithmic foundation for realizing IGRT-BNCT.

Keywords: ESSM-ML, IGRT-BNCT, Machine learning, Database, Computational efficiency

Introduction

As a binary targeted therapeutic modality, Boron Neutron Capture Therapy (BNCT) has emerged at the forefront of next-generation precision radiotherapy research in recent years [1, 2], demonstrating remarkably effective therapeutic outcomes for several cancer types, particularly head and neck cancer, malignant melanoma, and glioblastoma [3–5]. The fundamental mechanism of this technique hinges on the ^{10}B(n, α)^{7}Li nuclear reaction occurring between epithermal neutrons and boron-10 (^{10}B)-enriched tumor cells. This collision triggers the emission of α particles and lithium-7 (^{7}Li) ions with high linear energy transfer (LET). These particles exhibit traversal distances equivalent to a single cellular diameter (less than 10 µm), enabling selective destruction of tumor cells while maximizing preservation of adjacent normal tissues [6, 7]. During its early developmental phase, BNCT predominantly relied on nuclear reactors as neutron sources. However, the substantial size and inherent safety regulations associated with reactor-based systems have posed significant barriers to widespread clinical adoption within hospital settings [8, 9]. In recent years, with the advancement of compact accelerators, numerous countries including China, Japan, the United States, and Russia have initiated research programs aimed at developing accelerator-based BNCT systems (AB-BNCT) [10, 11].

The quality of the neutron beam critically influences the therapeutic efficacy of BNCT. The latest International Atomic Energy Agency (IAEA) report, *Advances in Boron Neutron Capture Therapy* [12], provides detailed design requirements for neutron beam parameters. Among these parameters, the neutron energy spectrum—which represents the distribution of neutron fluence rate or neutron density versus neutron energy—stands as one of the most critical factors [13–15]. Currently, the epithermal neutron fluence rate shall be $\geq 5 \times 10^{8}$ $\mathrm{cm}^{-2} \cdot \mathrm{s}^{-1}$, which is lower than the 1×10^{9} $\mathrm{cm}^{-2} \cdot \mathrm{s}^{-1}$ threshold recommended in IAEA-TECDOC-1223 (2001) [16]. Additionally, the thermal neutron fraction must be ≤ 0.05, and the fast neutron dose equivalent per unit epithermal neutron fluence shall not exceed 7×10^{-13} $\mathrm{Gy} \cdot \mathrm{cm}^{2}$.

For BNCT treatment planning systems (TPS), the neutron energy spectrum must be determined accurately to achieve proper therapeutic dosing, considering patient-specific parameters (weight, ^{10}B concentration in tumor), tumor characteristics (type, location, size), and critical organ dose-volume constraints [17–19]. Therefore, if the proton parameters of the accelerator can be dynamically adjusted in real-time according to tumor treatment response during radiotherapy, thereby modulating the incident neutron energy spectrum, enhanced tumor cell ablation can be achieved while sparing normal tissues—this concept is known as IGRT-BNCT. Compared with traditional input-output simulation methods using Monte Carlo that demand substantial computational time per calculation, the Nanjing University team proposed the Energy Spectrum Superposition Method (ESSM), which achieves significant computational efficiency enhancement by leveraging a pre-built Monte Carlo-based simulation database, reducing calculation time by a factor of 16,500 [19]. However, the ESSM method exhibits two limitations: (1) although the database is pre-computed, the construction process remains time-consuming; and (2) when encountering neutron or proton energies or directions not covered in the database, the uncertainty in the final energy spectrum results could be elevated.

Meanwhile, Machine Learning (ML), as a primary approach in artificial intelligence (AI), has been widely applied in radiation therapy and medical science [20–22]. As shown in Fig. 1, neural networks (NN) constitute a machine learning approach that mimics the brain's operational mechanism by referencing the connectivity structure of neurons; deep neural networks (DNN), also known as multi-layer perceptrons (MLP), are neural networks with multiple hidden layers [23, 24]. ML can achieve effective and rapid prediction of target data by training on extensive datasets [25]. Therefore, in this paper, two ML modules are integrated into the original ESSM master program, forming the ESSM-ML framework, to address these two limitations and to better implement IGRT-BNCT. Each ML module comprises two submodules: a DNN submodule and a random forest-based submodule [26], collectively aimed at predicting the required neutron spectra.

[Figure 1: see original paper]

II. Energy Spectrum Superposition Method-Machine Learning Coupling Algorithm

This study primarily includes the following components: (1) the theoretical foundation of ESSM and the ESSM-ML architecture, (2) database establishment, (3) neutron spectrum prediction and denormalization using DNN and RF, respectively, and (4) verification of the ESSM-ML results. Each component is described in detail below.

A. ESSM and ESSM-ML

1. ESSM Principle

The ESSM is a crucial approach for swiftly acquiring the neutron energy spectrum in BNCT, aiming to improve the computational efficiency of TPS and implement IGRT-BNCT. As shown in Fig. 2, it is grounded in the analysis of two distinct physical processes—proton-target interaction and neutron transport through the Beam Shaping Assembly (BSA)—with corresponding databases constructed for each process [19].

Proton-Target Interaction Process: (1) The proton spectrum and directional distribution from the accelerator are normalized. This normalization process is essential as it allows for the calculation of energy weights and direction weights, which represent the proportion of the fluence rate across different energy intervals and directional intervals, respectively. (2) Establish the Target Database. This database is populated with neutron energy spectra and neutron directional distribution information generated when single-energy proton beams strike targets from a single direction. (3) The neutron data corresponding to the specific proton energy spectrum and direction are retrieved from the Target Database. These data are then multiplied by their respective energy and direction weights. The sum of these weighted results yields the neutron energy spectrum generated by the interaction between the proton beam and the target. In practical calculations, considering the separation of energy and statistical processes, the integral form of the energy spectrum is converted into a summation form, and factors such as the proton-neutron conversion rate δ, target consumption rate L_{Target}, and other possible corrections K_{Target} are taken into account. The neutron energy spectrum and directionality formed by proton beams impacting targets are shown in Eq. (1):

The neutron energy spectrum and directional distribution ϕ*n(E,) E, after a single-energy, single-direction proton impacts the target:*

$$\phi_{Target}^{nK,nM}(E_{nk},\theta_{nm}) = K_{Target}L_{Target}\sum_{nI,nJ}\sum_{nK,nM}\delta(E_{pi},\theta_{nj})\phi_n^{pi,\theta_{nm}}(E_{nk})\phi_p(E_{ni},\theta_{nj})+\epsilon$$

where ϕ{Target} is the total neutron fluence rate from proton impacting target material, ϕ*{Target}(E*{nk}, _{nm}) is the neutron fluence rate with energy E_{nk} and direction _{nm}, E_{nk}^{pi} and _{nm} is the nk-th neutron energy and nm-th direction from proton with energy E_{pi} and direction _{pj}, represents other possible corrections that may exist, and it could also be in the form of a spectral distribution.

Neutron Entering the BSA Process: The processing approach for this process is similar to that employed for the proton-target interaction process. (1) Normalize the neutron energy spectrum and directional distribution resulting from the proton-target interaction as the energy and direction weights. (2) Establish the BSA Database. This database contains neutron energy spectra and neutron directional information calculated based on single-energy neutron beams with a single direction entering the BSA. (3) The relevant data are extracted from the BSA Database according to the target neutron parameters from the first physical process. These data are multiplied by the energy and direction weights obtained in step (1). The cumulative weighted results yield the neutron energy spectrum and directional distribution at the BSA emission window for this physical process. Eq. (2) shows the emission window spectrum from neutrons with energy c and direction _{nm} entering through the BSA:

$$\phi_n(E_{nk},\theta_{nm}) = \sum_{nP,nQ}\phi_n(E_{np}^{nk},\theta_{nq})$$

Therefore, integrating Eq. (1) and (2) and accounting for the proportion of neutrons lost α(E,) in the BSA due to decay, absorption, and scattering, along with other possible corrections s, the final emission window neutron energy spectrum for BNCT throughout the entire physical process is represented by Eq. (3):

$$\phi_{\text{window}} = \sum_{nK,nM}\phi_{\text{Target}}(E_{nk},\theta_{nm})[1-\alpha(E_{nk},\theta_{nm})]\phi_n(E_{nk},\theta_{nm}) + s$$

where ϕ_{window} is the total neutron fluence rate of the BSA emission window, and s could also be in the form of a spectral distribution.

2. ESSM-ML

As previously mentioned, to address the limitation of the ESSM algorithm stemming from missing data entries in the database, the method has been iteratively upgraded to ESSM-ML by leveraging the powerful generalization and rapid computational capabilities of machine learning. Fig. 3 [Figure 3: see original paper] shows the flowchart of the ESSM-ML calculation program. Two machine learning modules—ML Module of Target Database_{broad} and ML Module of BSA Database_{broad}—are integrated into the ESSM main program. For missing proton energy or angular data in the Target Database, the program will invoke pre-trained deep neural network models, DNN_{{{Target}}*{{database}}*}{broad} or DNN_{{{BSA}}*{{database}}*}{broad}, to predict the neutron energy spectra resulting from proton-target interactions corresponding to the missing energy or angle conditions. Subsequently, the neutron spectra predicted by the neural network undergo denormalization using the RF algorithm. The finalized spectra are then incorporated into the main program for subsequent computational analysis. If the database encompasses all data pertaining to the proton energy spectra and directional distributions generated by the accelerator, the program will bypass the ML module and proceed with subsequent execution. The invocation scenarios for the ML Module of BSA Database_{broad} are fundamentally analogous to the previous ML Module of Target Database_{broad} implementation, and thus no further elaboration will be provided here.

Moreover, a neutron dose calculation module has been integrated into the program to quantify the fast neutron absorbed dose contamination in the emerging neutron spectrum at the BSA emission window. The fast neutron absorbed dose D_f is given in Eq. (4):

$$D_f = \int_{10\mathrm{keV}}^{12\mathrm{MeV}} \phi(E)K(E)dE$$

where K, a function of E, is the neutron kerma coefficient $[\mathrm{Gy} \cdot \mathrm{cm}^2]$ for soft tissue as specified in the ICRU-63 Report [27, 28]. For missing energy points, log-log interpolation is employed in the same manner as in the Monte Carlo N-Particle Transport (MCNP) code. This interpolation scheme is appropriate for the ANSI/ANS flux-to-dose rate conversion factors, kermas for tissue, air and water, and energy absorption coefficients [29].

B. Simulation and Database Establishment

The structures of the target and the BSA, illustrated in Fig. 4 [Figure 4: see original paper] (a) and (b), are used to establish the Target Database and BSA Database. MCNP X-2.7.0 (Monte Carlo N-Particle) code, coupled with the ENDF/B-VII nuclear data library, is employed to simulate and acquire the required data [30].

1. Target and BSA

In this study, a 14 MeV cyclotron with a 1 mA proton current and a 5 cm proton beam radius is selected. As shown in Fig. 4, beryllium (Be) is chosen as the target material with an 8 cm radius and thickness, assuming a maximum proton penetration depth of 0.147 cm. Copper (Cu) and vanadium (V) are used as substrates to absorb the energy of the Bragg peak just behind the target. Downstream of the target, three layers of enriched uranium (^{235}U: 2.96% and ^{238}U: 97.04%) multipliers, each with a thickness of 0.05 cm, are added to enhance neutron production. The entire target assembly is enclosed within a Be sphere with a radius of 15 cm, which is then surrounded by a 5 cm-thick lead (Pb) spherical shell. The combination of U, Be sphere, and Pb spherical shell forms a moderating and reflective multiplier system that serves to enhance neutron yield while simultaneously reducing the energy of fast neutrons generated from proton bombardment [19].

The BSA is a crucial component of the Boron Neutron Capture Therapy (BNCT) facility, primarily used to moderate the fast neutrons generated from proton-target interactions and collimate them into neutrons suitable for therapeutic use [31–33]. The BSA features a three-layer moderator composed of AlF_3 (32.5 cm), CaF_2 (22.5 cm), and TiF_3 (2 cm). This combination maximizes epithermal neutron fluence while ensuring compliance with the latest IAEA report requirements. A 0.5 cm cadmium (Cd) layer and a 1 cm bismuth (Bi) layer are used to absorb thermal neutrons and gamma rays, respectively. The target structure and moderator structure are encased in a Pb cylinder shell with a length of 118.5 cm and a radius of 65 cm. The double-layer collimator, with inner Pb and outer polyethylene (PE) layers each 5 cm thick, focuses the neutron beam. The radius of the emission window is 8 cm.

Detailed descriptions regarding the structure and optimization of the target and BSA are provided in the article “The neutron energy spectrum obtained using energy spectrum superposition method for boron neutron capture therapy” [19]; therefore, no further elaboration will be given here. Table 1 presents a parameter comparison between this BSA and IAEA recommended values.

[Figure 4: see original paper]

2. Database Establishment

As mentioned above, the Target Database is related to protons interacting with the target. The proton energy ranges from 13 MeV to 14 MeV with an interval of 0.1 MeV, resulting in 11 energy points. The proton direction varies from -5° to +5° with an interval of 1°, totaling 11 directional points. As shown in Fig. 4(c), the neutron energy spectrum of each direction for three surfaces—the lateral sides (sideSurf), as well as the front (outSurf) and back (inpSurf) circular surfaces—of the cylindrical target is recorded. The recorded neutron energy of each spectrum spans from 10^{-8} MeV to 12 MeV: 40 logarithmic bins from 10^{-8} MeV to 5×10^{-7} MeV for thermal neutrons, 101 bins from 5×10^{-7} MeV to 10^{-2} MeV for epithermal neutrons in BNCT, and 72 bins from 10^{-2} MeV to 12 MeV for fast neutrons, totaling 214 energy points. The recorded neutron emission direction ranges from cos = -1 to cos = 1 with an interval of 0.05, giving 41 directional points. Moreover, due to the smaller radius of the proton beam compared to the Be target, for the outSurf and inpSurf, a circular area with a radius of 0.1 cm, one circular ring with a width of 0.4 cm, and fifteen circular rings each with a width of 0.5 cm are recorded (Fig. 4(c)). Ultimately, the Target Database comprises three sub-databases: Targ_{outSurf} database, Targ_{inpSurf} database, and Targ_{sideSurf} database.

The BSA database is for neutrons entering the BSA. For each ring of outSurf and inpSurf, the neutron direction varies from cos = 0 to cos = 1 with an interval of 0.05, resulting in 21 directional points. Since neutrons with energy below 0.01 MeV can barely penetrate the moderator, a total of 71 logarithmically equally-spaced energy points spanning from 0.01 MeV to 12 MeV for neutron energy are calculated. The neutron energy spectrum at the emission window is recorded. No database is established for sideSurf because the proportion of neutrons emitted from this surface accounts for less than 1%. The BSA Database comprises two sub-datasets: BSA_{outSurf} database and BSA_{inpSurf} database. The particle number for calculating each energy point and directional point is 10^7. The process of database establishment is also elaborated in detail in reference [19].

3. Data Preprocessing of Database

After database establishment, data preprocessing is required to meet the training requirements of DNN. The following processing steps were performed on the database:

(1) The proportion of neutrons emitted from the sideSurf is less than 1%, and the statistics of the energy spectrum in the Targ_{sideSurf} database are poor, which may mislead other training data of the DNN. Excluding this database does not affect the final results due to the extremely low proportion.

(2) In the Target Database, low neutron counts in certain energies, circular rings, directions, or energy regions of the spectrum, combined with some recording methods, result in partially noisy data or even negative values. First, energy spectra with extremely low proportions, poor statistics, and the potential to contaminate the training set were excluded. Then, a preprocessing algorithm was used to identify and process these noisy data. Table S1 shows the preprocessing algorithm, where {L*} represents the labels in the Target Database, and n_{ij} represents the neutron fluence rate at the j-th energy point of the i-th energy spectrum. No processing was needed for the BSA Database due to its good quality.

(3) The distribution and proportions of neutrons vary across different energy regions. Although BNCT primarily utilizes epithermal neutrons, the IAEA has established explicit requirements for both fast and thermal neutrons as well. Therefore, the energy spectrum is divided into thermal (10^{-8}–5×10^{-7} MeV), epithermal (5×10^{-7}–0.01 MeV), and fast neutron regions (0.01–12 MeV), with different DNN model parameters applied to each region.

(4) The energy spectrum data are normalized. For the Target Database, the features used for DNN and RF are input energy and direction of proton, square of circular ring, and output direction of target neutron. For the BSA Database, the features are input energy and direction of neutron. The labels are the neutron fluence rates corresponding to 214 energy points.

The t-SNE (t-distributed stochastic neighbor embedding) technique, a visualization method for compressing high-dimensional data into low-dimensional space, is used to visualize the results of data preprocessing on the Target Database [34]. To better intuitively show the difference between the original data and the preprocessed data of the Target Database, t-SNE is used to compress the 214-dimensional tensors into 2D (2-dimensional) and 1D (1-dimensional) tensors, respectively. The 2D scatter plots (Fig. 5 [Figure 5: see original paper] (A-a) and (B-a)) show that the original data and preprocessed points highly overlap and interweave into a “cluster-like” pattern. For the 1D histogram (Fig. 5 (A-b) and (B-b)) and 1D cumulative distribution function (CDF) (Fig. 5 (A-c) and (B-c)), the original data and preprocessed values also exhibit very high consistency within the same feature interval. The areas of the overlapping regions between the original data and the preprocessed data on the Inpsurf and Outsurf surfaces in the histogram reach 99.54% and 99.75%, respectively. These visualization graphs indicate that the distribution of the preprocessed data and the original data in the high-dimensional feature space is highly similar. The fluence rate distribution of the energy spectrum, after preprocessing such as normalization or elimination of abnormal spectra, remains unbroken and still retains the core physical features. However, the label distributions of the Targ_{outSurf} and Targ_{inpSurf} databases are different, which indicates that there are differences in the target neutron energy spectrum distributions between the two databases. Therefore, different network parameters will be used in subsequent DNN training and prediction.

[Figure 5: see original paper]

D. Machine Learning Model

1. Deep Neural Networks

As mentioned above, Deep Neural Networks (DNNs) are employed to predict the neutron energy spectrum. The basic building block refers to Multi-Layer Perceptron (MLP) and ResNet [35, 36]. Fig. 6 [Figure 6: see original paper] (a) shows a normal sequential connection block. To alleviate the gradient vanishing problem, increase network depth, and enhance feature reuse, a residual block with residual connections and a fully connected (FC) layer of dimension N^2 are introduced on this basis (Fig. 6 (b)). In addition, to avoid the gradient vanishing problem, the activation function ReLU was replaced with LeakyReLU (Eq. (5)) [37]. For the Target Database and the BSA Database, the DNN network parameters are respectively designed and optimized for different energy regions of the energy spectrum to achieve more accurate prediction results, namely DNN_{{{Target}}*{{database}}}*{broad} and

DNN_{{{BSA}}*{{database}}*}{broad}.

$$\text{LeakyReLU}(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases}$$

When training and testing the target neutron energy spectrum and the neutron energy spectrum from the emission window with MSE as the loss function, negative values often occur in the neutron fluence rate, and the fluence rate is zero at some energy points. This results in significant fluctuations in the overall neutron energy spectrum. Therefore, the Mask Non-Positive Loss function (MNPLoss), shown in Eq. (6), is designed to address these issues. MNPLoss consists of two components: MAELoss and a non-positive penalty. Additionally, to prevent overfitting and improve the model's generalization ability, an early stopping mechanism is introduced to save the best model at the triggering point [38]. When the early stopping mechanism is triggered, the best model and MNPLoss at that point are saved.

The results of the training and test sets for different databases are shown in Table S2 and Fig. S1 and Fig. S2. The MNPLoss curves of the training and validation sets basically converge to a stable and identical level as the number of epochs increases, although there is slight volatility for the two databases of the BSA Database. The MNPLoss values of the training and validation sets for the best model are of the same order of magnitude and relatively low. This indicates that the optimal model exhibits good performance without overfitting.

$$L_{MNP}(\text{pred}, \text{truth}) = L_{MAE}(\text{pred}, \text{truth}) + \lambda \cdot P(\text{pred})$$

$$L_{MAE}(\text{pred}, \text{truth}) = \sum \left| \text{pred} - y_{\text{truth}}^{(i)} \right|$$

$$P(\text{pred}) = \sum \left| y_{\text{pred}}^{(j)} \right| \quad \text{where } y_{\text{pred}}^{(j)} \leq 0$$

where λ is a hyperparameter to control the weight of the non-positive penalty; y_{truth}^(i) is the truth value; pred is the prediction; and y_{pred}^(j) is the non-positive prediction.

[Figure 6: see original paper]

2. Denormalization

The prediction result from the DNN is a normalized neutron energy spectrum. When this model was applied to predict the total neutron fluence rate, the prediction accuracy reached only 90%. Considering the model's generalization capability, traditional machine learning methods were subsequently employed for denormalization of neutron spectrum predictions. The outSurf dataset of the Target Database was utilized for preliminary evaluation of four models: Linear

Regression (LR), RF, Gradient Boosting (GB), and Support Vector Regression (SVR) [39–42]. To ensure an objective evaluation, the datasets were shuffled, and the average goodness-of-fit M-R^2 for 3-fold cross-validation and Root Mean Squared Error (RMSE) (Eq. (7)) were compared separately for these models. The RF model demonstrated the best performance with M-R^2 (0.998) and RMSE (5.89E-07) (Table S3).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

where n is the number of samples; y__i is the truth value; and $\hat{y}$__i is the predicted value.

Subsequently, the RF model was utilized to conduct further evaluations on four datasets from two databases. For each dataset, it was shuffled and then divided into a training set and a test set in an 8:2 ratio. Table 2 presents a comparison of M-R^2 and RMSE for the training set and test set across different databases. The relatively high M-R^2 and low RMSE indicate that the RF model exhibits strong performance in predicting the total neutron fluence rate. In addition, for the two sub-databases in the Target Database, the feature importance of Output Direction (emission direction of target neutrons) and Square (area of the circular ring) surpasses that of Input Energy (incident proton energy) and Input Direction (incident proton direction). For the two sub-databases in the BSA Database, the feature importance of Input Energy (incident neutron energy) surpasses that of Input Direction (incident neutron direction) (Fig. S3). Feature importance analysis visually illustrates the impact of different parameters on the total neutron fluence rate, which subsequently influences the final emission energy spectrum.

E. Verification of the ESSM-ML Results

The verification of ESSM-ML is primarily divided into three sections:

(1) For the proton-target interaction process, the ML Module of Target Database_{broad} is used to predict the neutron energy spectrum generated by 14 MeV protons hitting the target. These predicted spectra are then used to calculate the proportions of fast, epithermal, and thermal neutrons for different Output Directions, and these proportions are compared with the original data in the Target Database.

(2) For the neutron entering the BSA process, the ML Module of BSA Database_{broad} is used to predict the neutron energy spectrum at the emission window for neutrons with energies of 0.31392 MeV, 3.3359 MeV, and 8.0931 MeV from the BSA database. These predicted spectra are then used to calculate the proportions of fast, epithermal, and thermal neutrons, and these proportions are compared with the original data in the database. The three energy points are selected because they

span three orders of magnitude and are highly representative. For lower energies, the neutrons can barely penetrate the BSA.

(3) Both the traditional simulation-based method and ESSM are employed to calculate the neutron energy spectra at the emission window, with the proton energy spectrum set at 13 MeV, 13.2 MeV, 13.5 MeV, 13.7 MeV, and 14 MeV with proportions of 0.1, 0.2, 0.4, 0.2, and 0.1, respectively. The angular distribution of protons is set at 0°, 1°, 2°, 3°, 4°, and 5° with corresponding proportions of 0.6, 0.1, 0.1, 0.1, 0.05, and 0.05. Subsequently, the data entries from the database corresponding to 13.5 MeV and 14 MeV protons, as well as 0.31392 MeV, 3.3359 MeV, and 8.0931 MeV neutrons in all directions are removed, and the emission window neutron spectrum is then calculated using ESSM-ML. Finally, the result is compared with those obtained from the traditional simulation-based method and ESSM.

III. Results and Discussion

A. DNN Prediction Results

1. DNN of Target Database_{broad}

Since the DNN needs to predict more than 3,600 energy spectra, these spectra are superimposed and analyzed for convenient display and comparison. Fig. 7 [Figure 7: see original paper] shows the total neutron energy spectra (superposition of all input and output directions and circular ring energy spectra) from the outSurf and inpSurf surfaces for 14 MeV proton-target interaction.

In terms of spectral shape, the predicted energy spectrum is generally consistent with the ground truth, and the DNN correctly extracts both characteristic peaks (Peak1 and Peak2), although the Peak 2 characteristic on inpSurf is less distinct. For Peak1 on both surfaces, the predicted values are slightly higher than the truth.

The evaluation of the predicted target neutron energy spectra in different directions is provided in Table S4. As mentioned above, the radius of the target (R_{Target} = 8 cm) exceeds that of the proton beam (R_p = 5 cm). Therefore, to better present the results, the target is divided into two regions: 0 cm $\leq$ R $<$ 5 cm and 5 cm $\leq$ R $<$ 8 cm, with the prediction results for these two regions evaluated separately. For outSurf and inpSurf with 0 cm $\leq$ R $<$ 5 cm, the minimum R^2 values are 0.837 and 0.778, the maximum R^2 values are 0.946 and 0.968, and the mean R^2 values are 0.922 and 0.943, respectively. All RMSE values are less than 0.03, with mean values of 0.017 and 0.015, respectively. For outSurf and inpSurf with 5 cm $\leq$ R $<$ 8 cm, the minimum R^2 values are 0.490 and 0.494, and the maximum R^2 values are 0.779 and 0.778. All RMSE values are less than 0.043, with mean values of 0.038 and 0.039, respectively. Although the prediction results still show correlation with the truth values for the 5 cm $\leq$ R $<$ 8 cm region, the predictive performance for 0 cm $\leq$ R $<$ 5 cm is better. This may be because fewer target neutrons are produced in the region beyond the proton radius, with poorer statistics, thereby deteriorating the prediction results.

Better prediction results can be achieved by increasing the number of computational particles to improve the data quality of the database. Since establishing the database is a "sunk time cost" during the device's factory commissioning phase, this does not increase computational overhead in actual clinical practice. Overall, larger R^2 values and smaller RMSE values demonstrate that the DNN_{{{Target}}*{{database}}*}{broad} model has good predictive ability and accuracy for neutron energy spectra from the proton-target interaction process.

[Figure 7: see original paper]

2. DNN of BSA Database_{broad}

Fig. 8 [Figure 8: see original paper] shows the total energy spectra obtained by superimposing the emission window energy spectra of neutrons with different energies entering the BSA from various directions. In terms of spectral shape trends, the predictions are in good agreement with the truth values across three different energy levels. Since epithermal neutrons account for a large proportion (approximately 90%) of the energy spectrum and exhibit high statistical significance, the prediction values are very close to the truth values, with a maximum deviation of only 2.4% (8.0931 MeV in BSA_{outSurf}). For thermal neutrons, except for 21.8% at 0.31392 MeV in BSA_{inpSurf}, the deviations between prediction and truth values are within 15%, and for 3.3359 MeV and 8.0931 MeV, they are even below 5%. For fast neutrons, except for 26.2% at 3.3359 MeV in BSA_{inpSurf}, the deviations between predicted and true values are all within 20%. Although thermal and fast neutrons together account for a smaller proportion (approximately 10%) of the energy spectrum, resulting in poorer statistics, their overall prediction deviations remain below 20%, which is within an acceptable range.

The evaluation of the predicted neutron energy spectra in different directions is provided in Table S5 and Table S6. For the three energy points spanning different magnitudes across various databases, their R^2 values range from 0.970 to 0.981, and RMSE values range from 0.007 to 0.009. These extremely high R^2 and extremely low RMSE values further demonstrate the excellent predictive capability of the DNN_{{{BSA}}*{{database}}*}{broad} model for the energy spectra generated by the neutron entering the BSA process.

[Figure 8: see original paper]

B. Denormalization of Total Fluence Rate

Fig. S4 and Fig. S5 illustrate the prediction, based on RF, of total neutrons corresponding to protons or neutrons with specific energies in the Target Database and BSA Database. The R^2 of the prediction results are all greater than 0.99. Fig. 9 [Figure 9: see original paper] shows the relative deviations (RD) between prediction values and truth values in different directions, where RD = (prediction value - truth value) / truth value. As mentioned in Section 2.2, for the physical process of proton-target interaction, the target is divided into multiple circular rings, with emitted target neutrons recorded (Fig. 4(c)). For clear visualization, predictions from all circular rings in the same direction are summed to represent the total number of target neutrons emitted from the entire target in that direction, which are then compared with the truth values.

For the total neutron fluence rate corresponding to 14 MeV in the Target Database, the absolute RD values are generally less than 0.002, with a maximum value of -0.0053 for 14 MeV in Targ_{inpSurf} (cos( ) = 0.95). For the total neutron fluence rates corresponding to 0.31392 MeV, 3.3359 MeV, and 8.0931 MeV in the BSA Database, the absolute RD values are generally less than 0.02, with a maximum value of -0.047 for 3.3359 MeV in BSA_{outSurf} (cos( ) = 0.95). Higher R^2 and lower RD indicate that the prediction results can accurately denormalize the neutron energy spectra predicted by DNN.

[Figure 9: see original paper]

C. Comparison Among Traditional Simulation-Based Method, ESSM, and ESSM-ML

Fig. 10 [Figure 10: see original paper] and Table 3 present the neutron energy spectra obtained by different methods and the proportion of neutrons across various energy ranges. In terms of spectral shape, the three methods yield essentially identical results. Specifically, the energy spectrum of ESSM-ML is slightly lower than that of the traditional simulation-based method in both the epithermal and fast neutron energy regions. For neutron fluence rate fractions relative to the total, the epithermal neutron fraction of ESSM-ML is essentially consistent with that of ESSM and approximately 1.3% higher than that of the traditional method, while its fast neutron fraction matches that of ESSM and is roughly 1.5% lower than that of the traditional simulation-based method.

For the total neutron fluence rate, the ratios of ESSM-ML and ESSM relative to the traditional simulation-based method are 95.3% and 98.9%, respectively. For thermal and epithermal neutrons, the fluence rates of ESSM-ML are slightly lower than those of the traditional method (R_{thermal}^{ESSM}-ML/Trad = 101.75% and R_{epithermal}^{ESSM}-ML/Trad = 96.65%), while those of ESSM are essentially consistent with the traditional method (R_{thermal}^{ESSM}/Trad = 105.95% and R_{epithermal}^{ESSM}/Trad = 100.33%). For fast neutrons, R_{fast}^{ESSM}/Trad = 82.20%, which is approximately 20% lower than that of the traditional simulation-based method. The main reason for this is that the BSA Database has a smaller data volume compared to the Target Database. After fast neutrons are moderated by the BSA, fewer fast neutrons remain, leading to poor statistics, which degrades the data training performance and further impairs the fluence rate prediction in the fast neutron energy region. The fast neutron contamination D_f/ϕ_{epi} = 5.48×10^{-13} Gy $\cdot$ cm^2, representing an approximate 22.8% reduction compared to 7.1×10^{-13} Gy $\cdot$ cm^2 obtained by the traditional simulation-based method. This reduction is fairly close to the decrease ratio of the fast neutron fluence rate.

Moreover, the traditional simulation-based method calculates particle number as 5 × 10^{10} with a computation time T_{trad} of 5673 minutes, while the ESSM method only takes T_{ESSM} = 26 seconds with an AMD 7950x CPU. According to the particle number scaling principle of the ESSM method, this is equivalent to calculating a particle number of 4.629 × 10^{10} (10^7 × 5 × 6 + 1.5 × 10^7 × 73 × 21 × 2), representing a computation efficiency improvement of over 12,000 times. The ESSM-ML method has a computation time T_{ESSM}-ML of 69 seconds, achieving an efficiency enhancement of over 4,500 times due to its ESSM architecture. A total of 4080 + 4080 + 60 + 60 = 8,280 energy spectra were predicted by invoking DNN and RF, with a rough estimation of the unit energy spectrum computation time as (69 - 26)/8280 = 0.0052 seconds (GPU: RTX 1066s). Although the overall computation efficiency of ESSM-ML decreases compared to ESSM alone, the program's adaptability to different energy spectra and directional distributions is significantly improved, making it better suited to handle variations in clinical-grade radiotherapy practice.

Furthermore, the original database can be pre-expanded using the ML Model of Target Database_{broad} and ML Model of BSA Database_{broad} to reduce the energy and direction intervals and improve database accuracy. This would reduce the number of calls to the ML module, thereby decreasing computational time and enhancing efficiency. For instance, the time required to obtain energy spectra using only ESSM (without ML) is only 26 seconds.

[Figure 10: see original paper]

IV. Conclusion

This article systematically introduces the ESSM-ML method for rapidly obtaining the neutron spectrum of BNCT, with the following conclusions:

(1) ESSM reduces the energy spectrum acquisition time from the tens-of-minutes scale to the seconds scale by establishing and invoking databases. ESSM-ML further extends the upper application limit of ESSM by introducing a machine learning module, enabling rapid and accurate acquisition of neutron energy spectra required for treatment based on real-time changes in accelerator proton beam parameters. The integration of databases and machine learning provides a theoretical and algorithmic foundation for BNCT and other radiotherapy technologies based on IGRT.

(2) The ML Model of Target Database and ML Model of BSA Database consist of DNN and RF models, respectively dedicated to neutron energy spectrum prediction and spectrum denormalization. Both DNN and RF show good performance in predicting the target neutron energy spectra and total fluence rate generated by 14 MeV proton-target interaction, as well as the neutron energy spectra and total fluence rate at the emission window after neutrons of three distinct energy levels enter the BSA. Furthermore, the neutron energy spectrum obtained by ESSM-ML is essentially consistent with that obtained by traditional simulation-based methods, with the ratio of their total fluence rates being 95.3%.

(3) The prediction time for a single energy spectrum is extremely short. On one hand, it is possible to further improve the accuracy of existing databases and reduce the time cost of energy spectrum acquisition. On the other hand, with improved database accuracy, the program's frequency of invoking the ML module will decrease, further enhancing the speed of obtaining treatment energy spectra.

Acknowledgments

This work was supported by the Team Building Program of Nanjing University (No. 1480604114).

Author Contributions: All authors contributed to the study conception and design. Material preparation, data collection, simulation, calculation, and analysis were performed by Jun-Kai Yang, Hao-Peng Deng, Zhi-Meng Hu, Fang-Cong Zhang, Li-Kai Guo, Min Peng, Ping-Quan Wang, Hui Zhang, Bo-Wen Zhou, Chungming Paul Chu, and Giuseppe Gorini. The first draft of the manuscript was written by Jun-Kai Yang, Hao-Peng Deng, and Chungming Paul Chu, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

[1] H. Kumada, F. Naito, K. Hasegawa et al., Development of LINAC-based neutron source for boron neutron capture therapy in University of Tsukuba. *Plasma Fusion Res.* **13**, 2406006 (2018). doi: 10.1585/pfr.13.2406006

[2] Z.P. Qiao, Y.C. Hu, Q.-X. Jiang et al., Coin-structured tunable beam shaping assembly design for accelerator-based boron neutron capture therapy for tumors at different depths and sizes. *NUCL SCI TECH* **34**, 186 (2023). doi: 10.1007/s41365-023-01325-w

[3] S. Kawabata, M. Suzuki, K. Hirose et al., Accelerator-based BNCT for patients with recurrent glioblastoma: a multicenter phase II study. *Neuro-Oncol. Adv.* **3**(1), vdab067 (2021). doi: 10.1093/noajnl/vdab067

[4] J. Hiratsuka, N. Kamitani, R. Tanaka et al., Boron neutron capture therapy for vulvar melanoma and genital extramammary Paget's disease with curative responses. *Cancer Commun.* **38**(1), 38 (2018). doi: 10.1186/s40880-018-0297-9

[5] Z. Zhang, T. Liu, A review of the development of In-Hospital Neutron Irradiator-1 and boron neutron capture therapy clinical research on malignant melanoma. *Ther. Radiol. Oncol.* **2**, 49 (2018). doi: 10.21037/tro.2018.10.03

[6] Z.K. Liu, W.S. Cheng, W.-S. Wu et al., Principles and clinical application of boron neutron capture therapy. *Med. J. Peking Union Med. Coll. Hosp.* **14**, 698–705 (2023). doi: 10.12290/xhyxzz.2023-0120

[7] J. Li, Q. Sun, C. Lu et al., Boron encapsulated in a liposome can be used for combinational neutron capture therapy. *Nature* **13**, 2143 (2022). doi: 10.1038/s41467-022-29780-w

[8] R. L. Moss, F. Stecher-Rasmussen, K. Ravensberg et al., Design, construction and installation of an epithermal neutron beam for BNCT at the high flux reactor Petten. *Basic Life Sci.* **54**, 77–86 (1990). doi: 10.1007/978-1-4615-

3384-9_{11}
[9] H. Baba, Y. Onizuka, M. Nakao et al., Microdosimetric evaluation of the neutron field for BNCT at Kyoto University reactor by using the PHITS code. *Radiat. Prot. Dosim.* **143**(2–4), 528–532 (2011). doi: 10.1093/rpd/ncq511
[10] H. Tanaka, Y. Sakurai, M. Suzuki et al., Improvement of dose distribution in phantom by using epithermal neutron source based on the Be(p, n) reaction using a 30 MeV proton cyclotron accelerator. *Appl. Radiat. Isot.* **67**, S258–S261 (2009). doi: 10.1016/j.apradiso.2009.03.096
[11] A. J. Kreiner, J. Bergueiro, D. Cartelli et al., Present status of accelerator-based BNCT. *Rep. Pract. Oncol. Radiother.* **21**(2), 95–101 (2016). doi: 10.1016/j.rpor.2014.11.004
[12] International Atomic Energy Agency, *Advances in Boron Neutron Capture Therapy.* (IAEA, Vienna, 2023).
[13] L. Ren, Y.-C. Han, J.-C. Zhang et al., Neutronics analysis of a stacked structure for a subcritical system with LEU solution driven by a D-T neutron source for ^{99}Mo production. *NUCL SCI TECH* **32**(11), 123 (2021). doi: 10.1007/s41365-021-00968-x
[14] Y.N. Zhu, Z.K. Lin, H.Y. Yu et al., Study on the optimal incident proton energy of ^{7}Li(p,n)^{7}Be neutron source for boron neutron capture therapy. *NUCL SCI TECH* **35**, 60 (2024). doi: 10.1007/s41365-024-01420-6
[15] J.K. Yang, P.Q. Wang, Z.G. Ren et al., Comparison of neutron energy spectrum unfolding methods and evaluation of rationality criteria. *NUCL SCI TECH* **33**(12), 152 (2022). doi: 10.1007/s41365-022-01139-2
[16] IAEA-TECDOC-1223. *Current status of neutron capture therapy.* IAEA, 8, 75–77 (2001).
[17] G. Xiao, Z.-Q. Wang, B.-A. Zhang et al., A survey of BNCT treatment planning system. *Chin. J. Med. Phys.* **23**(1), 1–4 (2006). doi: 10.3969/j.issn.1005-202X.2006.01.002
[18] G. G. Daquino, Treatment planning systems for BNCT: requirements and peculiarities. In: *Proceedings of International Workshop on Neutron Capture Therapy*, Pisa University Press, Pisa, 45–52 (2003).
[19] J. Yang, H. Deng, P. Wang, The neutron energy spectrum obtained using energy spectrum superposition method for boron neutron capture therapy. *IEEE Trans. Nucl. Sci.* **72**(6), 1856–1863 (2025). doi: 10.1109/TNS.2025.3569730
[20] J. Zhang, Y. Zhang, Y. Yin et al., A review of machine learning in tumor radiotherapy. *J. Biomed. Eng.* **36**(5), 879–884 (2019). doi: 10.7507/1001-5515.201810051
[21] G. Valdes, C. B. Simone, J. Chen et al., Clinical decision support of radiotherapy treatment planning: a data-driven machine learning strategy for patient-specific dosimetric decision making. *Radiother. Oncol.* **125**(3), 392–397 (2017). doi: 10.1016/j.radonc.2017.10.014
[22] F. Yi, J. Yuan, F. Han et al., Machine learning reveals connections between preclinical type 2 diabetes subtypes and brain health. *Brain* **148**(3), awaf057 (2025). doi: 10.1093/brain/awaf057
[23] T. Meng, S. Shan, H. Shao et al., SE-GNN: expanded-aware graph neural network with iterative optimization for semi-supervised entity alignment.

IEEE Trans. Knowl. Data Eng. **37**(5), TKDE-2023-07-0426 (2025). doi: 10.1109/TKDE.2025.3555586

[24] G.L. Wang, K.-M. Chen, S.-W. Wang et al., Beam based alignment using a neural network. *NUCL SCI TECH* **35**, 75 (2024). doi: 10.1007/s41365-024-01436-y

[25] R. Zhang, Y. Yu, Z. Zhang et al., An iterative prediction method for designing the moderator used for the boron neutron capture therapy. *Med. Phys.* **49**(1), 598–610 (2021). doi: 10.1002/mp.15339

[26] L. Breiman, Random forests. *Mach. Learn.* **45**(1), 5–32 (2001). doi: 10.1023/A:1010933404324

[27] Y. Hashimoto, F. Hiraga, Y. Kiyanagi, Optimal moderator materials at various proton energies considering photon dose rate after irradiation for an accelerator-driven ^{9}Be(p,n) boron neutron capture therapy neutron source. *Appl. Radiat. Isot.* **106**, 88–91 (2015). doi: 10.1016/j.apradiso.2015.07.027

[28] ICRU, *Nuclear Data for Neutron and Proton Radiotherapy and for Radiation Protection.* (ICRU Report 63, Bethesda, 2000).

[29] M. R. James, D. B. Pelowitz, A. J. Fallgren et al., MCNP6™ User's Manual Code Version 6.1.1beta, Los Alamos National Laboratory Report LA-CP-14-00745, Los Alamos, (2014).

[30] J.S. Hendricks, G.W. McKinney, L.S. Waters et al, MCNPX Extensions Version 2.5.0. Los Alamos National Laboratory Tech. Rep., LA-UR-05-2675, (2005).

[31] J.Y. Chen, J.F. Tong, Z.-L. Hu et al., Evaluation of neutron beam characteristics for D-BNCT01 facility. *NUCL SCI TECH* **33**, 12 (2022). doi: 10.1007/s41365-022-00996-1

[32] B. A. Elshahat, A. Naqvi, K. Abdalla et al., Design calculations of an accelerator based BSA for BNCT of brain cancer. *J. Radioanal. Nucl. Chem.* **272**(2), 403–408 (2007). doi: 10.1007/s10967-006-6949-2

[33] X.Y. Wang, J.Y. Chen, Q. Zhang, Boron shielding design for neutron and gamma detectors of a pulsed neutron tool. *NUCL SCI TECH* **36**, 16 (2025). doi: 10.1007/s41365-024-01605-z

[34] L. van der Maaten, G. Hinton, Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(2605), 2579–2605 (2008). doi: 10.1145/1390156.1390267

[35] Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**(7553), 436–444 (2015). doi: 10.1038/nature14539

[36] K. He, X. Zhang, S. Ren et al., Deep residual learning for image recognition. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 770–778 (2016). doi: 10.1109/CVPR.2016.90

[37] J. Xu, Z. Li, B. Du et al., Reluplex made more practical: Leaky ReLU. In: *Proc. IEEE Symp. Comput. Commun.*, 1–7 (2020). doi: 10.1109/ISCC50000.2020.9219587

[38] S. Paguada, L. Batina, I. Buhan et al., Being patient and persistent: optimizing an early stopping strategy for deep learning in profiled attacks. *IEEE Trans. Comput.* **74**(3), 875–886 (2025). doi: 10.1109/TC.2023.3234205

[39] Scikit-learn: machine learning in Python. (Scikit-learn, 2025), https://scikit-learn.org/. Accessed 10 May 2025.

[40] W.D. Li, Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification. *Sens. Actuators B Chem.* **177**, 970–980 (2013). doi: 10.1016/j.snb.2012.11.071

[41] B.C. Wang, C.X. Tang, M.T. Qiu et al., A machine learning approach to TCAD model calibration for MOSFET. *NUCL SCI TECH* **34**, 192 (2023). doi: 10.1007/s41365-023-01340-x

[42] H. Liu, X.X. Li, Y. Yuan et al., Prediction of the first 2^+ states properties for atomic nuclei using light gradient boosting machine. *NUCL SCI TECH* **36**, 21 (2025). doi: 10.1007/s41365-024-01613-z

Note: Figure translations are in progress. See original paper for figures.