Prediction of Survival in Patients With Esophageal Cancer After Immunotherapy Based on Small-Size Follow-Up Data

Yuhan Su, Chaofeng Huang, Chen Yang, Qin Lin, and Zhong Chen

***Abstract*—Esophageal cancer (EC) poses a significant health concern, particularly among the elderly, warranting effective treatment strategies. While immunotherapy holds promise in activating the immune response against tumors, its specific impact and associated reactions in EC patients remain uncertain. Precise prognosis prediction becomes crucial for guiding appropriate interventions. This study, based on data from the First Affiliated Hospital of Xiamen University (January 2017 to May 2021), focuses on 113 EC patients undergoing immunotherapy. The primary objectives are to elucidate the effectiveness of immunotherapy in EC treatment and to introduce a stacking ensemble learning method for predicting the survival of EC patients who have undergone immunotherapy, in the context of small sample sizes, addressing the imperative of supporting clinical decision-making for healthcare professionals. Our method incorporates five sub-learners and one meta-learner. Leveraging optimal features from the training dataset, this approach achieved compelling accuracy (89.13%) and AUC (88.83%) in predicting three-year survival status, surpassing conventional techniques. The model proves efficient in guiding clinical decisions, especially in scenarios with small-size follow-up data.**

***Index Terms*—Esophageal carcinoma, immunotherapy, survival prediction, stacking ensemble learning, machine learning.**

***Impact Statement*—This study introduces a stacking ensemble learning model achieving 89.13% accuracy in predicting three-year survival for esophageal cancer patients receiving immunotherapy, offering valuable support for clinical decision-making.**

Received 10 January 2024; revised 5 February 2024, 9 July 2024, and 19 August 2024; accepted 26 August 2024. Date of publication 2 September 2024; date of current version 11 September 2024. This work was supported in part by the Natural Science Foundation of Xiamen City under Grant 3502Z20227007, in part by the Basic and Applied Basic Research Foundation of Guangdong Province under Grant 2022A1515110042, and in part by the Xiaomi Young Talents Program. The review of this article was arranged by Editor Yudong Zhang. *(Corresponding authors: Qin Lin; Zhong Chen.)*

Yuhan Su is with the School of Electronic Science and Engineering, Xiamen University, Xiamen 361005, China, and also with the Shenzhen Research Institute of Xiamen University, Shenzhen 518057, China.

Chaofeng Huang is with the Institute of Artificial Intelligence, Xiamen University, Xiamen 361005, China.

Chen Yang and Qin Lin are with the First Affiliated Hospital of Xiamen University, Xiamen 361000, China (e-mail: linqin05@163.com).

Zhong Chen is with the School of Electronic Science and Engineering, Xiamen University, Xiamen 361005, China, and also with the Institute of Artificial Intelligence, Xiamen University, Xiamen 361005, China (e-mail: chenz@xmu.edu.cn).

Digital Object Identifier 10.1109/OJEMB.2024.3452983

I. Introduction

ESOPHAGEAL cancer (EC) ranks ninth among the most prevalent malignant tumors worldwide, with the sixth-highest mortality rate attributed to cancer. The five-year survival rate of EC patients undergoing surgical intervention alone is low [1], [2], [3], [4], [5]. As the initial symptoms of EC are often atypical, patients commonly present in advanced stages, leading to an unfavorable prognosis [6], [7], [8]. In recent years, in addition to conventional therapeutic approaches such as surgery, radiotherapy, and chemotherapy, multidisciplinary neoadjuvant strategies like stent therapy, molecular targeted therapy, and immunotherapy have gained prominence in the management of EC. Immunotherapy, in particular, has emerged as a crucial adjunct to traditional modalities for EC treatment [9]. By potentially activating the patient's innate immune system, immunotherapy elicits a sustained anti-cancer response, aiming for durable disease control. Moreover, immunotherapy exhibits efficacy across various stages of EC, encompassing both advanced and pre-advanced stages. Notably, in comparison to certain traditional cancer treatments, immunotherapy typically presents with reduced adverse effects, contributing to an enhanced quality of life for patients [10]. Nevertheless, it remains essential to establish the optimal timing for immunotherapy, attain accurate prediction of the response to immunotherapy, identify the most appropriate cohort of patients for maximal benefits from immunotherapy, and optimize the survival advantage. As such, developing an accurate prognostic model for EC patients post-immunotherapy is crucial for enhancing prognosis and facilitating informed clinical decisions [11], [12], [13].

The realm of data mining methodologies stands poised to deliver substantial knowledge and insights, enabled by their capacity to systematically process vast troves of structured and unstructured data drawn from diverse sources. These methodologies possess the prowess to unveil patterns and synthesize robust predictive models using clinical data [14], [15], [16]. Notably, within this array of techniques, machine learning (ML) has ascended as a prominent paradigm, finding prolific application in forecasting cancer survival outcomes. Its effectiveness in augmenting prediction accuracy has been well documented [17]. Among the various ML methods, ensemble learning approaches

have garnered significant attention for their remarkable performance superiority over singular classifiers. These approaches can be broadly classified into two categories: homogeneous and heterogeneous ensemble learning methods. Of particular note is the stacking ensemble learning method [18], [19], [20], which hinges on harnessing heterogeneous weak learners. This involves integrating diverse learning algorithms and subsequently amalgamating the resultant base models through a meta-model. The efficacy of this strategy is particularly pronounced when dealing with limited sample sizes, such as in scenarios involving small datasets.

A. Related Works

Immunotherapy has been widely examined by experts and researchers. The authors in [11] evaluated the safety and feasibility of esophagectomy in patients with locally advanced EC who had received neoadjuvant immunotherapy and chemoradiotherapy. The results demonstrated that the combination of neoadjuvant immunotherapy and standard chemoradiotherapy followed by esophagectomy showed satisfactory outcomes in this patient population with locally advanced EC. [21] presented a rank-based pairwise comparison algorithm that was utilized for selecting effective immune-related gene pairs (IRGPs). The authors then proceeded to construct a prognostic IRGP signature utilizing the least absolute shrinkage and selection operator regression model. The immune signatures established have exhibited considerable potential in prognosticating the clinical outcome, tumor immunogenicity, and the response to immunotherapy in patients diagnosed with EC. In narrative review [22], the authors provided a overview of the current state of immunotherapy for EC, categorized by disease stage. They also discussed promising biomarkers and potential future directions in the field. The authors highlighted the ongoing need for additional prospective and dedicated clinical trials to identify molecular biomarkers capable of predicting treatment response and patient prognosis in the context of immunotherapy for EC. The prognostic implications of various indicators for patients with EC have been examined through distinct studies. For instance, [23] explored the prognostic relevance of lactate dehydrogenase (LDH) in the context of advanced EC patients undergoing immunotherapy. Similarly, in another investigation, the authors in [24] assessed the prognostic significance of the pretreatment platelet to albumin ratio among EC patients undergoing definitive radiotherapy. The authors in [25] employed a systematic review and meta-analysis approach to evaluate the safety and efficacy of neoadjuvant immunotherapy in EC. Results from the comprehensive analysis demonstrated the safety and efficacy of neoadjuvant immunotherapy in locally advanced EC patients, thereby warranting its consideration as a recommended neoadjuvant treatment modality for this population. Notwithstanding, the authors have emphasized the urgent need for additional research to reveal the long-term effects of immunotherapy, as emphasized in their work. This provides the impetus for our study.

The increasing availability of computer technology has facilitated the widespread use of artificial intelligence (AI) in the medical field. Among the AI techniques, ML has become a prominent research focus in medicine. Different ML approaches have demonstrated their effectiveness in predicting key factors, including tumor susceptibility, recurrence, and survival rates for malignancies. In recent years, numerous studies have employed ML techniques to model cancer risk and patient outcomes. For instance, a comprehensive review of the latest publications on the use of ML in cancer research was presented in [26]. In another study, [27] developed and validated a novel hybrid approach utilizing an improved synthetic minority oversampling technique and an adaptive support vector machine (SVM) method to predict the postoperative survival of lung cancer patients with imbalanced data. Ensemble learning methods are increasingly popular in the field of ML, as they are capable of enhancing the overall prediction performance of a model and addressing the issue of weak supervision. In the context of cancer research, numerous ensemble learning methods have been proposed to predict patient survival. For example, in [28], a homogeneous ensemble learning method was developed to predict the survival of breast cancer patients, using an area under the curve-based integration mechanism to combine twelve different SVM models. Results indicated that the proposed method outperformed SVM alone. Similarly, in [29], a two-stage tree ensemble-based model was proposed to predict the survival of colorectal cancer patients, which distinguished between patients whose outcome was survival or not and those whose outcome was death. However, the use of single models in homogeneous ensemble learning methods can lead to limitations due to the structural constraints of each model. Overcoming this limitation is challenging, and hinders improvement in model performance. To address this, the authors in [30] proposed a heterogeneous ensemble learning method to predict the survival of neuroblastoma patients and extract decision rules to aid physicians in making decisions. In another study, the authors in [31] introduced a parallel Bayesian hyperparameter optimized stacking ensemble model (referred to as BSense) designed for prognosticating breast cancer survival. This model was constructed through a stacking mechanism that combines multiple ML models. The process of hyperparameter tuning for these machine learning models was accomplished using Bayesian optimization with Gaussian Processes, resulting in the identification of optimal hyperparameters. The integration of parallelism and Bayesian optimization was strategically employed to mitigate the computational time associated with this process. Nonetheless, to overcome data imbalance or small sample size issues, effective integration mechanisms are necessary to improve prediction performance. This remains a significant challenge that requires further exploration in future studies.

B. Novelty and Contributions

In contrast to the majority of prior research that has primarily focused on predicting the survival of EC patients post-radiotherapy or surgery, this paper delves into prognosticating EC patient survival subsequent to immunotherapy-assisted radiotherapy or surgery. Notably, through a systematic and statistical analysis, this paper establishes a prediction model for generating effective predictions under small sample sizes. This model is designed to offer diagnostic guidance and technical support to address the challenges of predicting survival status outcomes in

EC patients treated with immunotherapy-assisted radiotherapy or surgery. Specifically, a stacking ensemble learning-based approach is presented, consisting of five heterogeneous sub-learners and a meta-learner, which first utilizes the training data to generate optimal feature subsets of each sub-learner, then the prediction results of the sub-learners are fed into the meta-learner for further prediction. To evaluate the performance of the proposed method, various mainstream ML (MML) methods, start-of-the-art (SOA) methods, and stacking ensemble learning mechanisms are compared and assessed using statistical and ML metrics. Additionally, valuable information is extracted from the prediction model to assist clinicians in decision-making. The main contributions of this paper are as follows:

- We present an analysis of clinical and follow-up data from real-world cohorts of EC patients who underwent immunotherapy. The primary objective is to assess the impact of different features on the three-year survival status of EC patients following immunotherapy. Through a thorough evaluation of the obtained data, we identify several features that exhibit a statistically significant association with the prognostic survival status of EC patients. These findings emphasize the potential predictive value of these features in determining clinical outcomes, while also providing a deeper understanding of the complex interplay between these indicators and their impact on clinical prognosis of EC patients undergoing immunotherapy.
- We propose a prediction method for determining the prognostic survival status of EC patients receiving immunotherapy. Our method utilizes stacking ensemble learning and leverages multiple features to improve prediction accuracy. Specifically, we use sub-learners that are trained on identical datasets but with varying feature subsets to optimize their prediction capabilities. The outputs of these sub-learners are then fed into a meta-learner, resulting in a refined and more accurate prediction model. Our research demonstrates the efficacy of our method in predicting three-year prognostic survival status outcomes for EC patients treated with immunotherapy, particularly in scenarios involving small sample sizes. Our findings highlight the potential of this approach to improve clinical decision-making and ultimately enhance patient outcomes.
- We assess the effectiveness of the proposed method for predicting the three-year prognostic survival status of EC patients undergoing immunotherapy, using real-world testing dataset. Our experimental findings demonstrate that leveraging stacking ensemble learning yields superior accuracy compared to mainstream and SOA methods. Additionally, the proposed model is efficient in predicting the three-year prognostic survival status of EC patients following immunotherapy. The results provide compelling evidence for the efficacy of our model and underscore its potential as a valuable tool for predicting clinical outcomes in EC patients receiving immunotherapy.

Overall, this paper's contributions lie in the development of a prediction model for EC patient survival following immunotherapy-assisted radiotherapy or surgery, which has potential clinical implications. The subsequent sections of this paper are structured as follows: Section II provides an overview of the dataset employed in this paper and presents a detailed description of the data preprocessing and feature analysis procedures that are employed. Section III outlines the training process for the proposed prediction method. In Section IV, we evaluate the performance of our prediction model using several evaluation metrics and present a discussion of the results. Finally, Sections V and VI offer a concluding summary of this paper and propose directions for future research, respectively.

II. Materials

A. Dataset

The precise prediction of patient survival outcomes in the context of immunotherapy represents a critical concern in contemporary prognostic cancer studies [33], [34], [35]. To address this challenge and improve the prognostic accuracy of survival outcomes for EC patients receiving immunotherapy, we conduct a retrospective analysis utilizing data collected from the First Affiliated Hospital of Xiamen University. Specifically, we conduct between January 2017 and May 2021, focused on 113 EC patients receiving immunotherapy, 75 of whom are newly diagnosed, while 38 are diagnosed with recurrent or metastatic EC. Through this analysis, we seek to investigate the efficacy of immunotherapy in the context of EC and assess its potential to improve patient survival outcomes. This study protocol received ethical approval from the Medical Ethics Committee of The First Affiliated Hospital of Xiamen University (No. XMYY-2023KYSB094). To ensure the confidentiality of patient information, individual identifiers were substituted with a distinct study identification number. As a result, the requirement for informed consent was thus exempted.

The dataset captured comprehensive patient information, including general information, disease information, immunotherapy information, and prognosis follow-up information of the EC patients with immunotherapy. Specifically, the general information includes patients' EC type, age, gender, and personal history of other malignancies within five years (PHM), while disease information includes Eastern Cooperative Oncology Group (ECOG) score, interval from the time of the first diagnosis of EC to the time of follow-up (TFD), primary tumor site (PTS), metastatic situation (MS), pathology (PSCC), disease stage (DS), presence of oligometastatic disease (POD), pre-treatment diet (PTD), presence of supportive treatment (PST), and blood index information. The immunotherapy information includes immunotherapy drugs (ID), immunotherapy cycles (IC), presence of immune delay (PID), presence of local treatment (PLT), local treatment range (LTR), and local treatment time (LTT). The prognostic follow-up information consists of presence of disease progression (PDP) and three-year survival status, where the participants were followed by various methods, including telephone interviews, surveillance at the hospital, and medical reports in the inpatient department after recurrence. The primary outcome was ascertained by two investigators blinded to the predictor variables. Notably, the analysis did not include data during immunotherapy, such as the blood index information during immunotherapy, as the purpose of this paper is to provide a reliable prediction model to assist doctors in making

TABLE I
DETAILS AND INTERPRETATION OF FEATURES

Definition	Abbreviation	Measurement	Range	Explanation
General information				
EC type of the patients	Type	Binary	$\{0,1\}$	0: Newly diagnosed, 1: Recurrent and metastatic
Age of the EC patients	Age	Years	[45-91]	Chronological age count
Gender of the EC patients	Gender	Binary	$\{0,1\}$	0: Male, 1: Female
Personal history of other malignancies within five years	PHM	Binary	$\{0,1\}$	0: Patients without malignancies history 1: Patients with malignancies history
Disease information				
Eastern Cooperative Oncology Group	ECOG	Standard	$\{0,1,2,3,4,5\}$	Refer to score standard in [32]
Time interval of the EC patients	TFD	Month	[24-36]	Interval from the time of the first diagnosis of EC to the time of follow-up
Primary tumor site the EC patients	PTS	Index	$\{1,2,3\}$	1: Upper chest, 2: Mid chest, 3: Lower chest
Metastatic situation the EC patients	MS	Binary	$\{0,1\}$	0: No metastasis, 1: Metastasis was found
Pathology (squamous cell carcinoma)	PSCC	Index	$\{0,1,2,3\}$	0: Undifferentiated, 1: Poorly, 2: Moderately, 3: Well
Disease stage the EC patients	DS	Index	$\{1,2,3,4\}$	1: I-stage, 2: II-stage, 3: III-stage, 4: IV-stage
Presence of oligometastatic disease	POD	Binary	$\{0,1\}$	0: No oligometastatic, 1: Oligometastatic
Pre-treatment diet the EC patients	PTD	Index	$\{1,2\}$	1: Semi-liquid, 2: Normal diet
Presence of supportive treatment	PST	Binary	$\{0,1\}$	0: No supportive treatment, 1: Supportive treatment
Baseline white blood cell count	BWBC	$10^9/L$	RI: [3.5-9.5]	Actual detection value count of blood routine
Baseline neutrophil count	BNC	$10^9/L$	RI: [1.5-8.0]	Actual detection value count of blood routine
Baseline lymphocyte count	BLC	$10^9/L$	RI: [0.8-4.0]	Actual detection value count of blood routine
Baseline platelet count	BPC	$10^9/L$	RI: [100-300]	Actual detection value count of blood routine
Baseline albumin level	BA	g/L	RI: [35-50]	Actual detection value count of blood routine
Baseline lactate dehydrogenase	BLDH	U/L	RI: [140-280]	Actual detection value count of blood routine
Immunotherapy information				
Immunotherapy drugs of the EC patients	ID	Index	$\{1,2\}$	1: Camrelizumab, 2: Pembrolizumab
Immunotherapy cycles of the EC patients	IC	Count	[1-17]	21 days as a immunotherapy cycle
Presence of immune delay	PID	Binary	$\{0,1\}$	0: No immune delay, 1: With immune delay
Presence of local treatment	PLT	Binary	$\{0,1\}$	0: No local treatment, 1: With local treatment
Local treatment range of the EC patients	LTR	Index	$\{0,1,2\}$	0: No local treatment, 1: Partial lesion, 2: All lesions
Local treatment time of the EC patients	LTT	Index	$\{1,2,3\}$	1: Upper chest, 2: Mid chest, 3: Lower chest
Prognosis follow-up information				
Presence of disease progression	PDP	Binary	$\{0,1\}$	0: No disease progression, 1: With disease progression
Surviving status of the EC patients	Status	Binary	$\{0,1\}$	0: Died within three years, 1: Survival at follow-up

a diagnosis of patients before participating in immunotherapy. For ease of reference, the full names, corresponding abbreviations and explanations of the key information in this paper are listed in Table I. Note that blood index information lacks a precise range, prompting the provision of reference intervals (RI). Additionally, when evaluating the PLT, both surgery and radiotherapy were considered if the patient had undergone such interventions.

B. Data Preprocessing and Feature Analysis

This section details the crucial data preprocessing and feature analysis steps employed in the mining of immunotherapy prognosis data for EC patients. Initially, features exhibiting strong correlation with prediction outcomes are eliminated. For instance, PDP, closely tied to both immunotherapy prognosis and survival status, is precluded as a feature. For features with missing values, we applied the following methods: (i) removal of variables with missing rates exceeding 50% from the initial dataset; (ii) estimation of corresponding values of features with missing rates between 50% and 10% using the mean; and (iii) removal of patients with missing rates of features below 10% from the initial dataset. After preprocessing, the dataset consisted of 21 features and 112 samples. Secondly, to eliminate inconsistencies arising from features of different dimensions, we employed the minimum-maximum normalization method to normalize the dataset before constructing the prediction model [36]. Specifically, the z-th feature of the n-th sample can be transformed using the following formula:

$$x'_n(z) = \frac{x_n(z) - \min_n x_n(z)}{\max_n x_n(z) - \min_n x_n(z)}, \tag{1}$$

where $x_n(z)$ represents the initial value for the z-th feature of the n-th sample and $x'_n(z)$ is the converted value for the z-th feature of the n-th sample.

Following data preprocessing, it is imperative to investigate the interdependence among different features. The Pearson correlation coefficient (PCC) [37], a commonly used statistical measure to assess the strength of a linear relationship between two variables, can be utilized to determine the correlation between the features. In this, the PCC $\mathcal{P}$ between z-th feature and z'-th feature can be defined as follows.

$$\mathcal{P} = \frac{\sum_{n=1}^{N}(x'_n(z) - \overline{x'(z)})(x'_n(z') - \overline{x'(z')})}{\sqrt{\sum_{n=1}^{N}(x'_n(z) - \overline{x'(z)})^2}\sqrt{\sum_{n=1}^{N}(x'_n(z') - \overline{x'(z')})^2}}, \tag{2}$$

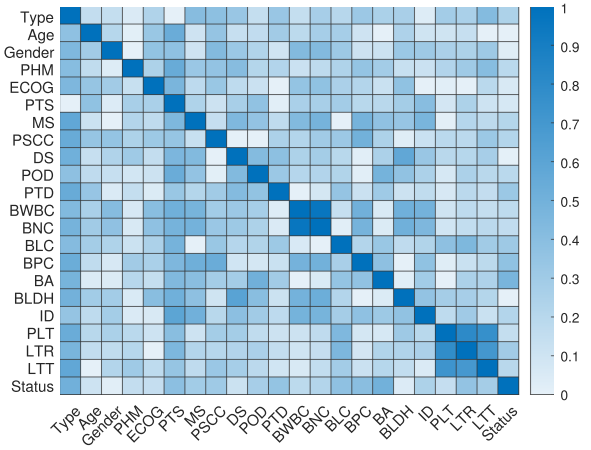


Fig. 1. Correlation heat map between the features. A heat map showing the value of the correlation coefficient between each possible pair of features.

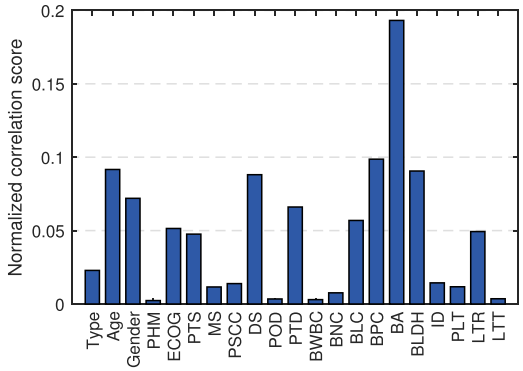


Fig. 2. Normalized correlation scores between the features and survival status.

where $\overline{x'(z)}$ and $\overline{x'(z')}$ are the means of two features, and N denotes the number of the samples. Upon conducting a correlation analysis on the preprocessed data, this paper employs the PCC to measure the relationship between features and presents a correlation coefficient heat map in Fig. 1 to visually represent these relationships. The heat map uses colors to indicate the degree of correlation, with darker or lighter colors indicating higher correlation. The results of the analysis reveal that the majority of the variables exhibit correlation coefficients less than 0.5, with most coefficients being less than 0.3.

C. Feature Extraction

Feature extraction plays a vital role in reducing model complexity and enhancing accuracy by eliminating redundant or irrelevant features from high-dimensional data [38]. It encompasses two main methods: filtering and encapsulation. The filtering method assigns scores to each feature based on divergence or correlation, applies a threshold or selects a specific number of features, and efficiently selects the most relevant ones. On the other hand, the encapsulation method explores and optimizes the feature space, iteratively selecting or excluding features based on an objective function, typically a prediction performance score. In this paper, a filtering feature extraction method is employed. Specifically, as the objective is to predict the survival status of EC patients following immunotherapy, the SelectKBest feature analysis method is utilized for each feature [39]. This method calculates the relative correlation of each feature by considering the highest correlation score with the survival status. The preprocessed dataset of EC patients is analyzed to determine the significant relationships between each feature and three-year survival status. The normalized correlation scores derived from this analysis are visually represented in Fig. 2, revealing that BA, BPC, DS, and age demonstrate the strongest correlations, in sequential order, with the prediction of three-year survival outcomes in EC patients following immunotherapy. These findings align with the experimental results reported in [23], [24], [40]. Additionally, Fig. 2 also illustrates a significant correlation between LTR in immunotherapy information and Status, which further supports the conclusion drawn in [41]. Specifically, older age is associated with poorer prognosis; patients with advanced tumors typically have a worse prognosis; a strong immune response predicts better outcomes; patients with superior BA and BPC have a better prognosis; and comprehensive treatment regimens contribute to improved prognosis. These findings align with existing biological knowledge and clinical experience, further validating the soundness of our analytical approach.

Following feature extraction, the top 11 features with the highest correlation scores are retained, including Age, Gender, ECOG, PTS, DS, PTD, BLC, BPC, BA, BLDH, and LTR. This set of 11 features is established as the final indicator system and used as the feature dataset for training the prediction model. The dataset is split randomly, with 70% used for training and the remaining 30% for testing the model's performance.

III. Methods

To bolster the precision of survival outcomes among EC patients who have undergone immunotherapy, we introduce an approach grounded in stacking ensemble learning. By harnessing a repository of real-world data originating from post-immunotherapy EC patients, we aim to enhance the predictive prowess for determining survival status. Fig. 3 delineates the framework of our proposed method, designed to prognosticate the survival trajectory of EC patients subsequent to immunotherapeutic interventions. The crux of our approach lies in the orchestration of a learning paradigm that amalgamates a spectrum of ML methodologies. By amalgamating diverse ML techniques through a cascade system, this approach effectively distills salient insights from the dataset, unfurls prognostic patterns from comparatively diminutive sample sizes, and furnishes steadfast prognoses concerning the three-year survival trajectories of EC patients subjected to immunotherapy regimens. The underpinning architecture of the proposed methodology encompasses a sub-learning module synergistically collaborating with a meta-learning module to synergistically achieve a heightened degree of prognostic precision.

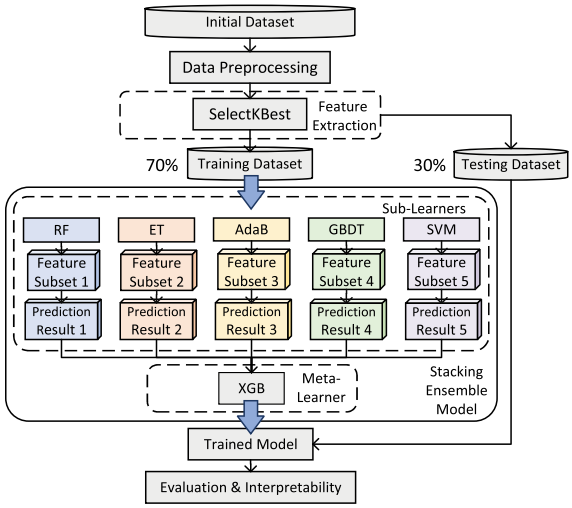


Fig. 3. Framework of the proposed prediction method.

A. Sub-Learning Module

From the perspective of stacking ensemble learning, the utilization of sub-learners can provide feature subsets that demonstrate optimal performance. However, it has been observed that the use of the same feature subset for guiding the construction of a stacking ensemble learning model may lead to relatively suboptimal ensemble performance. To achieve a high-quality ensemble, it is necessary to obtain individualized feature subsets for each sub-learner prior to constructing the prediction model. Thus, the optimal feature subset for each sub-learner can be used as prior knowledge to guide the construction of the prediction model. This paper has employed a sub-learning module consisting of five commonly utilized ML algorithms, namely random forest (RF) [42], extremely randomized trees (ET) [43], adaptive boosting (AdaB) [44], gradient boosting decision trees (GBDT) [45], and SVM [46], as sub-learners.

1) RF: RF is a supervised learning technique that has found applications in both classification and regression analyses. With a reputation for high accuracy, it is a flexible and straightforward algorithm that can be applied to a variety of datasets. A forest in RF comprises multiple trees that are constructed by bootstrapping the training data, and for each split, a randomly selected subset of features is employed. A higher number of trees in the forest translates to increased robustness. RF utilizes a decision tree approach, where data is randomly chosen to create multiple trees, and prediction is obtained from each tree. The best solution is then determined using a voting technique.

2) ET: ET is an ensemble classification technique based on decision trees. This approach integrates a higher level of randomness than the RT algorithm. The fundamental principle of ET is to use all the initial data for each decision tree and randomly select split nodes, with the goal of reducing variance more effectively than methods that employ weaker randomization schemes. The specific algorithmic steps of ET are as follows: (i) In the ET classification model, all samples are used for training in each decision tree. (ii) To increase randomness, features are randomly selected for each node split. The optimal attribute is chosen for each node to split the node, and this step is repeated until a decision tree is produced. (iii) The process outlined in (i) and (ii) is repeated to create an ET classification model.

3) AdaB: AdaB, a widely used and highly effective ensemble learning technique, is a meta-algorithm that belongs to the family of Boosting methods. Boosting is a variation of the bagging ensemble approach that improves learners by focusing on the areas where the system is underperforming. AdaB aims to build multiple models of the same classifier, with each one learning to correct the prediction errors of the previous model in the sequence. The individual classifier outputs are then combined into a weighted sum, which represents the final output of the boosted classifiers. This boosting technique is referred to as adaptive since subsequent weak learners are adjusted in favor of the misclassified instances. In AdaB, instances in the dataset are weighted according to their level of difficulty in being classified, enabling the algorithm to allocate more attention to these instances when building subsequent models. The algorithm assigns a higher weight to the misclassified instances, and each subsequent boosting iteration learns a new classifier on the weighted dataset. The classifiers are then weighted to form a single powerful classifier, with those having a low training error rate receiving a higher weight. The process is terminated using cross-validation.

4) GBDT: GBDT is a ML algorithm that has demonstrated remarkable success in various real-world applications. GBDT is employed for function estimation and is regarded as a method for numerical optimization in the function space. Specifically, GBDT relies on boosting to construct a robust model by iteratively adding weak learners that minimize a specified loss function. In each iteration, the weak learner is trained on the gradient of the loss function with respect to the current model's output. The resulting model is a weighted sum of the weak learners, with the weights determined by their performance in minimizing the loss function. Through this approach, GBDT progressively improves the performance of the model until a stopping criterion is met.

5) SVM: SVM is a statistical learning technique widely used for classification, regression, and outlier detection. It has been extensively applied in data analysis, especially in classification and regression tasks. In regression, SVM utilizes a nonlinear mapping function to transform low-dimensional samples into a high-dimensional vector space, which addresses the challenge of small-sample problems and enhances the efficacy of the method. In our study, SVM is employed as a sub-learner for predictive analysis, serving as a discriminant classifier that is defined by distinct hyperplanes. The algorithm operates on labeled training data, utilizing a supervised learning approach to obtain an optimal hyperplane for classifying novel instances.

To optimize the utilization of the available data, this paper employs a 10-fold cross-validation strategy for each sub-learner, as illustrated in Fig. 4. The dataset is partitioned into K subsets, with one subset allocated for validation and the remaining for

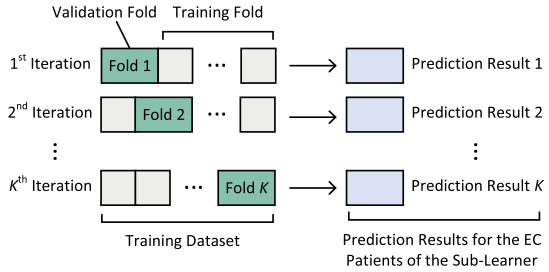


Fig. 4. Framework of the K-fold cross-validation approach for each sub-learner.

training during each of the K iterations. This process guarantees that all data undergo training K times and are validated once. Consequently, each sub-model's prediction results constitute a fusion of K validation results.

The sub-learner ensemble comprises a regression-based model and four tree-based models, as observed. These sub-learners are trained on the same dataset, enabling the determination of feature importance associated with the four decision tree-based classification algorithms. The computation of feature importance is based on the reduction of impurity at each node and its subsequent weighting to calculate the probability of reaching that node, commonly referred to as the Gini importance [47]. It quantifies the importance of each feature by summing the number of splits across all decision trees that incorporate the feature, weighted by the number of samples it splits. The probability of a node is derived by dividing the number of observations that reach the node by the total number of observations. A higher feature value signifies greater importance in the feature importance methods. Specifically, assuming that the dataset has a total of I classes, in a decision tree, each feature corresponds to a node, then the Gini coefficient $\mathcal{G}_z$ of node z is:

$$\mathcal{G}_z = 1 - \sum_{i=1}^{I} \left(\frac{N_z^i}{N_z} \right), \tag{3}$$

where N_z represents the number of samples of node z, and N_z^i represents the number of samples of class i of node z. Then, the evaluation of Gini Importance of node z in the decision tree is conducted as follows:

$$I_z = \frac{N_z}{N} \left(\mathcal{G}_z - \frac{N_z^{\mathrm{R}}}{N} \mathcal{G}_z^{\mathrm{R}} - \frac{N_z^{\mathrm{L}}}{N} \mathcal{G}_z^{\mathrm{L}} \right), \tag{4}$$

where $\mathcal{G}_z^{\mathrm{L}}$ and $\mathcal{G}_z^{\mathrm{R}}$ are the Gini coefficients of the left child and right child of node z respectively, N_z^{L} and N_z^{R} are the number of the samples of node z that go to the left child and go to the right child respectively. Feature importance plays a pivotal role in predictive modeling by providing valuable insights into the data, the model, and facilitating dimensionality reduction and feature selection to enhance the efficiency and effectiveness of prediction models. In this paper, the importance of each feature to the survival status is defined as the exponential function value of the Gini importance. Fig. 5 illustrates the evaluation of feature importance for each tree-based sub-learner. Notably, the feature importances vary across the training models associated with different ML algorithms. To leverage this observation, the proposed approach incorporates the prediction results of the five sub-learners as new feature subsets into the meta-learner. This novel prediction model combines the features from all five sub-learners, thus further reinforcing the accuracy of the predictions.

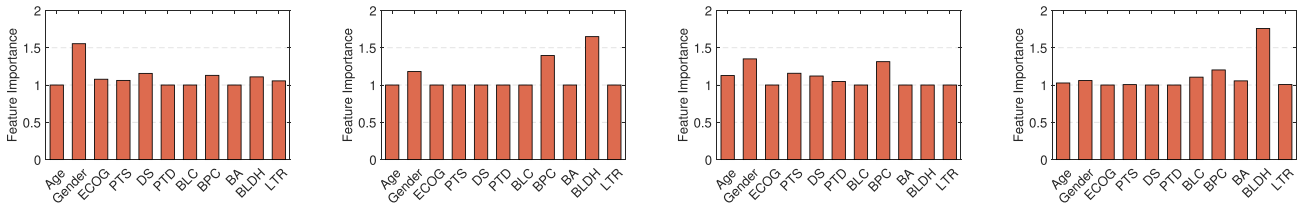
B. Meta-Learning Module

Single prediction models have inherent limitations in terms of their prediction performance. To overcome these limitations, ensemble learning techniques have been widely employed to integrate the strengths of individual models and construct superior prediction models. Extensive research has demonstrated the outstanding prediction performance of stacking ensemble learning [20], [48], [49], [50]. In the proposed prediction method, stacking ensemble learning is employed to enhance the performance of the prediction model by amalgamating multiple models. Specifically, the prediction results obtained from the sub-learning module, which constitutes the first layer prediction model, serve as input features for the second layer prediction model, known as the meta-learning module. The meta-learner leverages these features to learn the underlying relationships and subsequently generates the desired prediction results, as illustrated in Fig. 6. Extreme gradient boosting (XGB), a ML algorithm renowned for prediction and classification tasks, employs posterior inference to construct robust classifiers and regression models [51]. By effectively reducing model parameters using second-order Taylor expansion, XGB enhances evaluation accuracy and runtime efficiency. Notably, XGB exhibits several advantages, including simplicity, user-friendliness, and strong robustness. Hence, this paper employs XGB as the algorithm for the meta-learner.

In the meta-learner, let x_n^m denote the prediction result of sub-learner n for sample n, vector $\mathbf{x}_n = [x_n^1, \ldots, x_n^m, \ldots, x_n^M]$, vector $\mathbf{y}_n$ denote the associated label, and $\mathcal{D} = \{(\mathbf{x}_n, \mathbf{y}_n)\}$ denote the input dataset of the meta-learner, where $|\mathcal{D}| = N$. The final prediction function is given by

$$F_B(\mathbf{x}_n) = \sum_{b=1}^{B} f_b(\mathbf{x}_n) = F_{B-1}(\mathbf{x}_n) + f_B(\mathbf{x}_n), \tag{5}$$

where $f_b(\mathbf{x}_n)$ represents the model of the b-th base classifier for $\mathbf{x}_n$, $F_{B-1}(\mathbf{x}_n)$ represents the aggregation of predictions from the first $B-1$ base classifiers that have already been trained and fixed. On the other hand, $F_B(\mathbf{x}_n)$ corresponds to the final prediction made by the meta-learner, which combines the predicted values of the B base classifiers trained on $\mathbf{x}_n$. The objective function of the meta-learner consists of two components: a loss function and a regularization penalty term. The inclusion of the regularization penalty is essential to prevent overfitting and enhance the generalization ability of the model. The explicit mathematical form of the objective function is provided as



(a) ET (b) AdaB (c) RF (d) GBDT

Fig. 5. Feature importance of the different tree-based models in the sub-learning module.

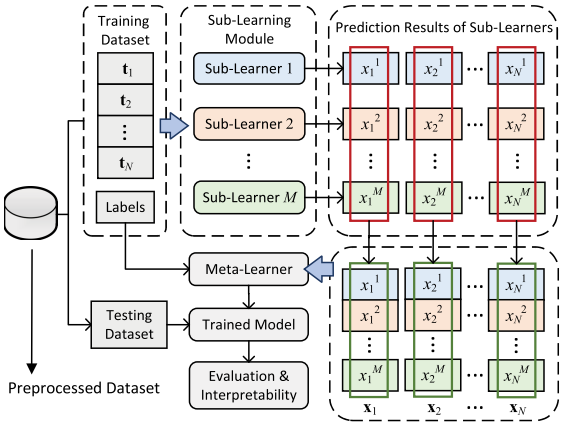


Fig. 6. Framework of the meta-learning module.

follows:

$$L_B = \sum_{n=1}^{N} \text{loss}\left[\mathbf{y}_n, F_B(\mathbf{x}_n)\right] + \sum_{b=1}^{B} \Omega(f_b)$$
$$= \sum_{n=1}^{N} \text{loss}\left[\mathbf{y}_n, F_{B-1}(\mathbf{x}_n) + f_B(\mathbf{x}_n)\right] + \sum_{b=1}^{B} \Omega(f_b). \quad (6)$$

The loss function, denoted as loss[.], evaluates the performance of a single sample prediction and quantifies the discrepancy between the predicted and actual values. It is assumed to be a convex function. On the other hand, $\Omega(.)$ represents the regularization term, which characterizes the complexity of the model. The regularization term is formulated as follows, capturing the contribution of model complexity to the overall objective:

$$\Omega(f_b) = \gamma J + \frac{1}{2}\lambda \sum_{j=1}^{J} \omega_j^2, \quad (7)$$

where J represents the number of nodes in the base classifier f_b, while ω_j denotes the weight of node j. Additionally, λ and γ are penalty coefficients incorporated in the formulation. As a consequence, when the complexity of the base classifier increases, the objective function is augmented, thereby effectively mitigating the risk of overfitting.

By applying the second-order Taylor expansion to the objective function (6) around $F_{B-1}(\mathbf{x}_n)$, an expanded function can be derived as follows:

$$L_B \cong \sum_{n=1}^{N} \Big[\text{loss}\left[\mathbf{y}_n, F_{B-1}(\mathbf{x}_n)\right] + g_n f_B(\mathbf{x}_n) + \frac{1}{2} h_n f_B^2(\mathbf{x}_n)\Big] + \sum_{b=1}^{B} \Omega(f_b), \quad (8)$$

where g_n and h_n are the first-order and second-order derivatives of the loss function at $F_{B-1}(\mathbf{x}_n)$, respectively, and are expressed as follows:

$$\begin{cases} g_n = \dfrac{\partial \text{loss}\left[\mathbf{y}_n, F_{B-1}(\mathbf{x}_n) + f_B(\mathbf{x}_n)\right]}{\partial F_{B-1}(\mathbf{x}_n)}, \\ h_n = \dfrac{\partial^2 \text{loss}\left[\mathbf{y}_n, F_{B-1}(\mathbf{x}_n) + f_B(\mathbf{x}_n)\right]}{\partial^2 F_{B-1}(\mathbf{x}_n)}. \end{cases} \quad (9)$$

Given that the first $B-1$ base classifiers have already been determined, the terms associated with $f_B(\mathbf{x}_n)$ in (8) become constants and do not impact the final optimization solution. Hence, we can eliminate the constant term, resulting in the transformation of the objective function as follows:

$$\tilde{L_B} = \sum_{n=1}^{N} \left[g_n f_B(\mathbf{x}_n) + \frac{1}{2} h_n f_B^2(\mathbf{x}_n)\right] + \Omega(f_B). \quad (10)$$

Based on (10), it can be inferred that once $F_{B-1}(\mathbf{x}_n)$ is determined, the values of g_n and h_n can be readily computed for each sample n.

$$\tilde{L_B} = \sum_{n=1}^{N} \left[g_n f_B(\mathbf{x}_n) + \frac{1}{2} h_n f_B^2(\mathbf{x}_n)\right] + \gamma J + \frac{1}{2}\lambda \sum_{j=1}^{J} \omega_j^2$$
$$= \sum_{j=1}^{J} \left[\left(\sum_{n \in \mathcal{N}_j} g_n\right) \omega_j + \frac{1}{2}\left(\sum_{n \in \mathcal{N}_j} h_n + \lambda\right) \omega_j^2\right] + \gamma J, \quad (11)$$

where $\mathcal{N}_j = \{\mathbf{x}_n | q(\mathbf{x}_n) = j\}$ represents the sample set on node j, $q(\mathbf{x}_n)$ represents the index function that maps samples to nodes, $\omega_j = f_B(\mathbf{x}_n)$ $(n \in \mathcal{N}_j)$ represents the regression value of node j. Let us define:

$$\begin{cases} G_j \triangleq \sum_{n \in \mathcal{N}_j} g_n, \\ H_j \triangleq \sum_{n \in \mathcal{N}_j} h_n. \end{cases} \quad (12)$$

By doing so, the objective function can be reformulated as a one-dimensional quadratic function with respect to the node weight

ω_j. To minimize the objective function, we can set its derivative to zero and solve for the optimal prediction score of each node as follows:

$$\omega_j^* = -\frac{G_j}{H_j + \lambda}. \tag{13}$$

Bringing (11) into the objective function, the minimum loss is obtained as:

$$L_B^* = -\frac{1}{2}\sum_{j=1}^{J}\left(\frac{G_j^2}{H_j + \lambda}\right) + \gamma J. \tag{14}$$

Then the meta-learner uses the following formula to evaluate the pros and cons of node splitting:

$$\text{Gain} = \frac{1}{2}\left[\frac{G_{\text{L}}^2}{H_{\text{L}} + \lambda} + \frac{G_{\text{R}}^2}{H_{\text{R}} + \lambda} - \frac{(G_{\text{L}} + G_{\text{R}})^2}{H_{\text{L}} + H_{\text{R}} + \lambda}\right] - \gamma, \tag{15}$$

where G_{L}, H_{L}, G_{R}, and H_{R} indicate the scores associated with different sub-trees following a split operation, respectively. The structural score Gain signifies the disparity between the loss prior to splitting and the loss subsequent to splitting. A greater disparity denotes a reduced loss after splitting, resulting in a lower value for the objective function, thus indicating improved effectiveness.

At this stage, the training of the meta-learner is complete, resulting in the acquisition of the final prediction model. Subsequently, the test dataset is employed to assess and evaluate the performance of the prediction model.

IV. Experiment

The Python language, along with the Scikit-learn libraries, was utilized to develop and evaluate the proposed prediction method using a range of ML techniques [52]. To ensure a fair comparison, all the models were implemented with their default hyperparameters and stopping criteria. To evaluate the effectiveness and superiority of the proposed method, we conducted three distinct experiments: (i) the comparison with MML-based models, (ii) the comparison with SOA method-based prediction models, (iii) the comparison with various stacking ensemble learning mechanism-based prediction models. For each model, the following metrics are evaluated on the dataset.

1) Accuracy: The accuracy metric in a prediction model quantifies the percentage of correctly predicted samples relative to the total samples, providing a measure of the model's overall effectiveness and precision. Nonetheless, it is important to acknowledge that this metric can be influenced by the distribution of samples across different categories. In scenarios where the number of samples in one category is significantly lower, such as the case where only 1% of samples correspond to the deceased category, even a model with limited discriminatory power may yield a high accuracy rate of 99%.

2) Precision: Precision serves as a prevalent evaluation metric within prediction models, encapsulating the accuracy of predicted survival outcomes in this paper. Specifically, it quantifies the ratio of accurately predicted occurrences where EC patients, specifically, those who succumbed to mortality within a three-year duration post-immunotherapy, are correctly identified among the total predicted death cases.

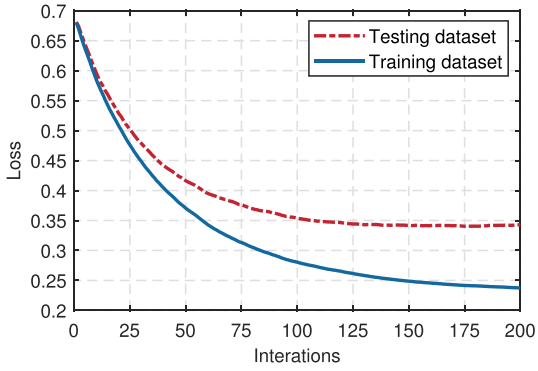


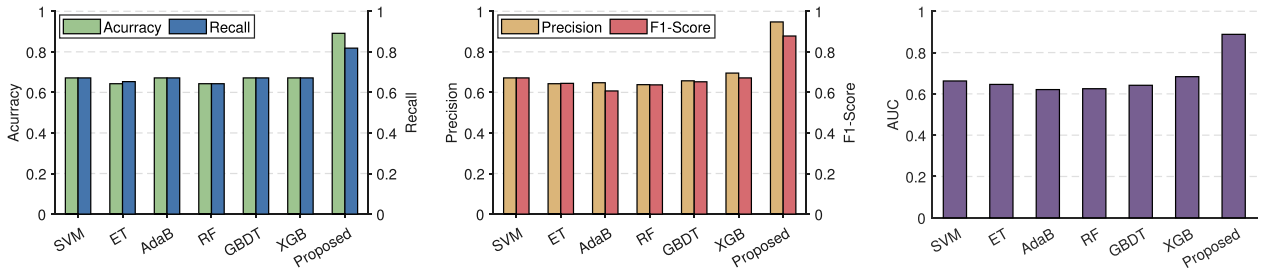
Fig. 7. Loss function on traning and texting datasets of the proposed prediction model.

3) Recall: The recall metric, a significant measure in evaluating prediction models, quantifies the ratio of actual positive samples (i.e., EC patients who experienced mortality within three years after immunotherapy in this paper) to the total positive samples encompassing both true positives and false negatives within a given dataset. Widely utilized to assess a model's recall or sensitivity, this metric holds utmost importance in assessing the model's efficacy in correctly identifying positive instances.

4) F1-Score: The F1-score, a performance metric in machine learning, effectively captures the intricate relationship between precision and recall. It offers a balanced evaluation of model performance, eliminating the need for explicit knowledge of the overall sample count. As a valuable alternative to conventional accuracy measures, the F1-score uncovers nuanced insights about the model that might be concealed when focusing solely on precision or recall optimization. By considering both precision and recall, the F1-score establishes a meaningful standard for evaluating model accuracy.

5) Area Under Curve (AUC): AUC is a commonly employed metric for evaluating prediction models, which quantifies the area under the receiver operating characteristic (ROC) curve. This paper adopts AUC as an evaluation criterion due to the inherent ambiguity of the ROC curve in determining the superior model. By providing a unified measure of model performance, AUC enables a more straightforward comparison of predictive efficacy, with larger AUC values generally indicative of higher predictive capability. Notably, AUC is computed as the summation of areas under the ROC curve, ensuring a comprehensive assessment of model performance across the entire spectrum of potential thresholds.

The iteration process of the proposed model is depicted in Fig. 7. It is evident that the model's loss steadily decreases with increasing iteration cycles until it reaches a stable value. Notably, the rate of decrease in loss for the test set is lower compared to that of the training set. Once the loss function attains a stable value, the loss function for the testing dataset consistently remains higher than that of the training dataset.



(a) Accuracy and Recall (b) Precision and F1-score (c) AUC

Fig. 8. Performance comparison of the mainstream traditional machine leaning-based prediction models.

TABLE II
PERFORMANCE COMPARISON OF THE DIFFERENT MAINSTREAM MACHINE LEARNING-BASED PREDICTION MODELS

Model	Accuracy	Recall	Precision	F1	AUC
SVM	0.6714	0.6714	0.6714	0.6714	0.6625
ET	0.6428	0.6531	0.6428	0.6446	0.6458
AdaB	0.6714	0.6714	0.6476	0.6071	0.6208
RF	0.6428	0.6428	0.6381	0.6371	0.6250
GBDT	0.6714	0.6714	0.6571	0.6523	0.6416
XGB	0.6714	0.6714	0.6952	0.6714	0.6833
Proposed	**0.8913**	**0.8182**	**0.9474**	**0.8780**	**0.8883**

A. Comparison With Mainstream Machine Learning-Based Prediction Models

We commence by comparing the presented prediction model with several MML-based models, namely SVM, ET, AdaB, RF, GBDT, and XGB. Fig. 8 displays the experimental results. Among the MML-based models, the performance parity among distinct models is a prevailing observation. Notably, the results underscore the significant superiority of the proposed model over the MML-based models. Detailed analysis, as presented in Table II, reveals that the proposed model outperforms the best-performing MML-based model (XGB) across various metrics. Specifically, the proposed model demonstrates 32.7% higher accuracy, 21.8% higher recall, 36.2% higher precision, 30.7% F1-score, and 30.0% higher AUC. Overall, the performance of the proposed model surpasses that of the other models. This can be attributed to the utilization of stacking ensemble learning in the proposed prediction method, which effectively integrates multiple prediction models with substantial differences. By leveraging the collective decision-making of the sub-learners, the proposed approach mitigates the impact of incorrect predictions by individual sub-learners, thus enhancing the overall prediction accuracy.

B. Comparison With Start-of-The-Art Method-Based Prediction Models

As discussed in Section I-B of this paper, the field of medical science has embraced the growing utility of ML algorithms, particularly in the context of survival prediction. This domain has witnessed notable contributions such as the prognosis of breast cancer patients' outcomes and the determination of the survival status of individuals diagnosed with neuroblastoma. A spectrum of endeavors has been undertaken to construct diverse prediction models, leveraging distinct datasets and a range of ML methodologies. However, it's essential to recognize that the applicability of these techniques varies substantially, and their effectiveness across distinct datasets might not be uniform. To ascertain the efficacy of the novel approach introduced in this study for predicting the survival of patients affected by EC post-immunotherapy, this research undertakes a comprehensive comparative analysis. Specifically, it conducts a juxtaposition between our proposed model and four SOA method-based prediction models: self-organizing maps (SOM)-SVM [53], XGBLC [54], DRGXG [30], and BSense [31]. The particulars of the comparison methods are delineated as follows:

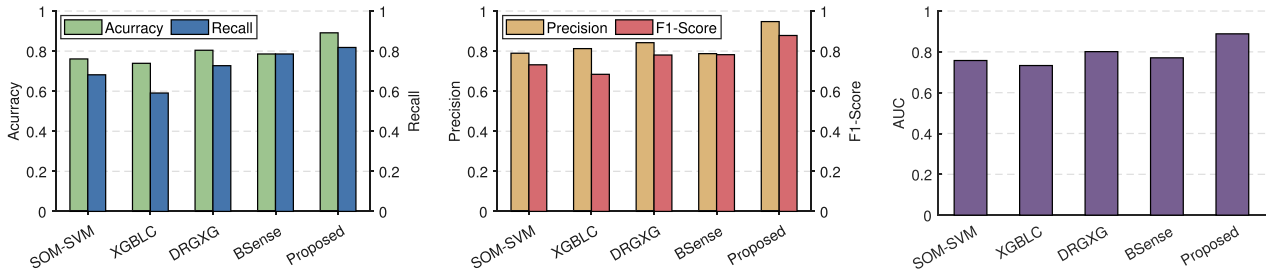
1) SOM-SVM: This method is proposed by the authors in [53], which is a prediction approach employing SOM neural network clustering and SVM techniques to prognosticate the survival risk levels in the context of EC.

2) XGBLC: This method is introduced by the authors in [54], which is a refined framework termed XGBLC, enriching the survival prediction model based on XGB through the incorporation of Lasso-Cox for a more robust analysis of high-dimensional genomic data.

3) DRGXG: This method is presented by the authors in [30], which is a heterogeneous ensemble learning scheme that takes the stage to predict survival in the realm of neuroblastoma. Notably, sub-learners in this method encompass decision trees, RF, a genetic algorithm-fueled SVM, XGB, and light gradient boosting machine (GBM).

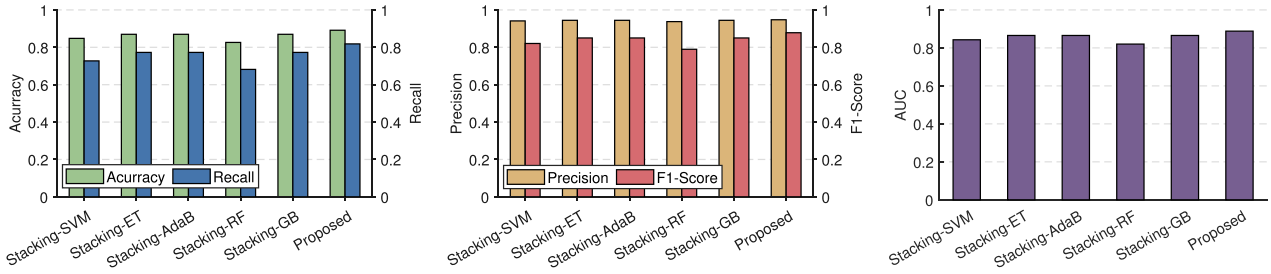
4) BSense: This method is presented by the authors in [31], which is a novel stacking ensemble prediction paradigm based on Bayesian hyperparameter optimization. This method, specifically designed for breast cancer survival prediction, employs a stack of sub-learners including deep neural networks (DNN), GBM, and distributed RF, and DNN also serves as the meta-learner.

Fig. 9 portrays the outcomes of our experimental investigations, accentuating the juxtaposition between the SOA method-based prediction model and the approach put forth in this paper. The outcomes distinctly exhibit the excellence of the proposed model. For instance, the proposed model attains an AUC of 0.8883, surpassing the SOM-SVM, XGRL, DRGXG,



(a) Accuracy and Recall (b) Precision and F1-score (c) AUC

Fig. 9. Performance comparison of the different traditional single machine leaning-based prediction models.



(a) Accuracy and Recall (b) Precision and F1-score (c) AUC

Fig. 10. Performance comparison of the different stacking ensemble learning mechanism-based prediction models.

TABLE III
PERFORMANCE COMPARISON OF THE DIFFERENT SOA METHOD-BASED PREDICTION MODELS

Model	Accuracy	Recall	Precision	F1	AUC
SOM-SVM [53]	0.7609	0.6818	0.7895	0.7317	0.7576
XGBLC [54]	0.7391	0.5909	0.8125	0.6842	0.7330
DRGXG [30]	0.8043	0.7273	0.8421	0.7805	0.8011
BSense [31]	0.7857	0.7857	0.7873	0.7822	0.7708
Proposed	**0.8913**	**0.8182**	**0.9474**	**0.8780**	**0.8883**

TABLE IV
STACKING ENSEMBLE LEARNING MECHANISM-BASED PREDICTION MODELS AS BENCHMARKS

Model	Sub-learners	Meta-learner
Stacking-SVM	ET, AdaB, RF, GBDT, XGB	SVM
Stacking-ET	SVM, AdaB, RF, GBDT, XGB	ET
Stacking-AdaB	SVM, ET, RF, GBDT, XGB	AdaB
Stacking-RF	SVM, ET, AdaB, GBDT, XGB	RF
Stacking-GB	SVM, ET, AdaB, RF, XGB	GBDT

and BSense methods by margins of 17.2%, 21.1%, 10.8%, and 15.2%, respectively. Moreover, the results, as outlined in Table III, corroborate the preeminence of the proposed model in forecasting the survival status of EC patients post-immunotherapy across four additional metrics. This is because the disparity in efficacy emerges as the SOM-SVM and XGRL methods are founded upon a single model predicated on substantial sample size. Notably, even the XGRL method enhances the performance of the XGB framework, it relies on a dataset of over 10,000 cancer patient samples to attain favorable prediction outcomes. As a consequence, its suitability for scenarios involving small sample sizes is limited. While the DRGXG method employs a ensemble learning architecture, its efficacy just lies in the amalgamation weighted by the sub-model AUCs, not re-optimization sub-learner outcomes. The BSense method, akin to our proposed method, employs a cascade framework to make secondary predictions based on sub-learner outcomes. However, the inherent limitation surfaces in its reliance upon a trifecta of sub-models, only three sub-learners hinder its predictive ability when faced with limited sample size.

C. Comparison With Stacking Ensemble Learning Mechanism-Based Prediction Models

To further validate the efficacy of the proposed model, we established five distinct stacking ensemble learning mechanism-based prediction models as benchmarks for comparison, as presented in Table IV. Fig. 10 depicts the experimental results, emphasizing the comparison between the various stacking ensemble learning mechanism-based prediction models and the proposed model. The outcomes reveal that the proposed model outperforms the benchmarks. For instance, the accuracy value achieved by the proposed model is 0.8913, surpassing the performance of Stacking-SVM, Stacking-ET, Stacking-AdaB, Stacking-RF, and Stacking-GB models by 5.1%, 2.5%, 2.5%, 10.4%, and 2.5%, respectively. Furthermore, the results for

TABLE V
PERFORMANCE COMPARISON OF THE DIFFERENT STACKING ENSEMBLE LEARNING MECHANISM-BASED PREDICTION MODELS

Model	Accuracy	Recall	Precision	F1	AUC
Stacking-SVM	0.8478	0.7273	0.9412	0.8205	0.8428
Stacking-ET	0.8696	0.7727	0.9444	0.8500	0.8655
Stacking-AdaB	0.8696	0.7727	0.9444	0.8500	0.8655
Stacking-RF	0.8261	0.6818	0.9375	0.7895	0.8201
Stacking-GB	0.8696	0.7727	0.9444	0.8500	0.8655
Proposed	**0.8913**	**0.8182**	**0.9474**	**0.8780**	**0.8883**

the remaining four metrics, as displayed in Table V, further substantiate the superior predictive capabilities of the proposed model in determining the post-immunotherapy survival status of EC patients. This superiority can be attributed to the utilization of XGB as the meta-learner in the proposed model. XGB leverages an optimized second-order Taylor expansion, allowing for better fitting of complex nonlinear datasets using second-order functions. Moreover, XGB effectively reduces model parameters, leading to more accurate predictions and faster execution. Additionally, XGB demonstrates superior handling of missing data and achieves more efficient completion of prediction tasks.

V. CONCLUSION AND DISCUSSION

EC is a prevalent tumor affecting the digestive tract, which presents a formidable hurdle in accurately prognosticating the survival outcomes of patients following immunotherapy, thus accentuating the pivotal nature of this task in tumor prognosis. In this paper, we introduced a novel approach grounded in the principles of stacking ensemble learning, aiming to construct a prediction model of exceptional quality for prognosticating the survival status of EC patients subsequent to immunotherapy. The proposed methodology encompasses a series of pivotal steps. Primarily, leveraging preprocessed data, we discerned the optimal feature sub-set specific to each sub-learner. These feature sub-sets were subsequently harnessed to construct the sub-learning module. Consecutively, the prediction outcomes derived from the sub-learners assume the role of essential prior knowledge in the construction of the meta-learning module. The central objective of this research is to holistically assess the performance exhibited by the proposed stacking ensemble learning prediction model and duly substantiate the efficacy of the proposed methodology. By meticulously scrutinizing the experimental results, a host of significant conclusions have been gleaned, succinctly summarized as follows:

- In the proposed heterogeneous stacking ensemble learning approach, the utilization of distinct sub-learners necessitates the selection of feature subsets that exhibit optimal performance for each sub-learner. This process guides the subsequent ensemble construction and enhances the overall integration effectiveness. In comparison to prediction techniques reliant on individual ML classifiers such as SVM, ET, AdaB, RF, GB, and XGB, commonly employed in various cancer prognosis and diagnosis investigations, the proposed method demonstrates effective prediction capabilities even under conditions of limited sample sizes.
- Through meticulous assessments and comparisons with diverse stacking ensemble learning methodologies encompassing five heterogeneous sub-learners, the presented stacking ensemble learning mechanism establishes its supremacy in terms of prediction performance, as substantiated by multiple ML metrics. Particularly, the proposed mechanism harnesses the inherent strengths of XGB's posterior inference to forge a resilient meta-learning module and classifier, resulting in augmented evaluation accuracy, accelerated execution speed, and heightened resilience.
- Through an extensive examination of the feature variables, our comprehensive study has yielded noteworthy insights. Firstly, within the cohort of EC patients receiving immunotherapy, age, gender, and DS have demonstrated a robust association with prognostic survival. This observation underscores the importance of these variables in the context of survival prognosis for EC patients undergoing immunotherapy. Secondly, pre-treatment blood testing has emerged as a pivotal determinant in the immunotherapeutic landscape for EC patients, wherein the values of BA, BPC, BLDH and LTR have exhibited a notable impact on survival prognosis. This is consistent with the research findings in [23], [24], [40], [41]. Consequently, healthcare professionals are advised to consider these findings diligently during the assessment of immunotherapy suitability for EC patients before initiating treatment.

Moreover, the rationale behind this study lies in the exclusive use of clinical features for survival prediction, providing notable advantages, particularly in terms of model simplification and heightened interpretability:

1) Model Simplification: Clinical features, encompassing fundamental physiological and pathological patient information, obviate the need for intricate imaging analyses. This results in the development of survival prediction models that are succinct, facile to construct, and straightforward to interpret.

2) Ease of Acquisition: Clinical features are typically derived from routine medical examinations and standard laboratory tests. In contrast, certain baseline imaging characteristics may necessitate costly imaging technologies that might not be universally available across medical institutions.

3) Interpretability: Clinical features often possess intuitive meanings, fostering easier comprehension and engendering trust among physicians and researchers regarding the predictive outcomes derived from these features.

VI. LIMITATIONS AND FUTURE DIRECTIONS

Given that the dataset utilized in this paper originates solely from a single medical institution, with a small sample size primarily composed of East Asian populations, the model's generalization across different genetic backgrounds (e.g., Europe, Africa, or America), and diverse patient groups remains unverified, potentially impacting its applicability and predictive performance. To further authenticate the validity and versatility of our model, forthcoming studies will undertake the following measures:

1) Multi-Center and Diversified Data Collection: We will collaborate with medical institutions in various regions and countries to gather data from patients with different genetic backgrounds. This multi-center approach will facilitate the evaluation of the model's predictive performance and applicability across diverse genetic backgrounds.

2) External Validation: By incorporating independent datasets from varied genetic backgrounds for external validation, we can comprehensively assess the model's generalization capability and reliability.

3) Model Adjustment and Optimization: Based on the data from populations with different genetic backgrounds, we will make necessary adjustments and optimizations to the model to enhance its prediction accuracy and applicability across diverse populations.

Moreover, it is imperative to acknowledge that relying solely on clinical features may sometimes fall short of providing a comprehensive patient profile. Certain imaging baseline features could encompass more nuanced and intricate information. In specific scenarios, amalgamating clinical and imaging features might enhance the model's predictive efficacy, enabling a more precise prognosis of immunotherapy outcomes. Consequently, the consideration of baseline imaging characteristics hinges on the specific objectives of the study and the characteristics of the available data. This consideration forms part of our prospective research agenda.

Furthermore, future investigations can also explore various aspects to further enhance our understanding. Firstly, integrating clinical data with RNA datasets of EC patients will enable the development of a holistic survival prediction model, potentially yielding valuable insights and refining prediction accuracy. Secondly, expanding the predictive methodology to encompass other tumor types, such as renal tumors and osteosarcoma, promises to advance prognostic research in oncology. Evaluating the proposed method's efficacy across diverse tumor types will furnish valuable insights for clinical application. Lastly, given the growing interest in heterogeneous ensemble learning techniques, investigating the performance of deep learning models in cancer survival prediction presents a promising avenue for future inquiry. Exploring the potential of deep learning approaches holds the potential to augment prediction accuracy and facilitate informed decision-making within clinical settings.

REFERENCES

[1] H. Sung et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] Z. Lin et al., "CT-guided survival prediction of esophageal cancer," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2660–2669, Jun. 2021.

[3] V. M. Merabishvili, "State of the russian oncology service: Esophageal cancer. Population-based study at the federal district level. Part II. Median survival, observed and relative survival of patients considering disease stage and histological structure of tumors," *Pharm. Formulas*, vol. 3, no. 3, pp. 40–47, 2021.

[4] P. Fang et al., "Prognostic value of micro-RNA 375, 133, 143, 145 in esophageal carcinoma: A systematic review and meta-analysis," *Front. Oncol.*, vol. 12, 2022, Art. no. 828339.

[5] E. J. Seibel et al., "Tethered capsule endoscopy, a low-cost and high-performance alternative technology for the screening of esophageal cancer and Barrett's esophagus," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1032–1042, Mar. 2008.

[6] J. C. Layke and P. P. Lopez, "Esophageal cancer: A review and update," *Amer. Fam. Physician*, vol. 73, no. 12, pp. 2187–2194, 2006.

[7] A. F. Hassanabad, R. Chehade, D. Breadner, and J. Raphael, "Esophageal carcinoma: Towards targeted therapies," *Cellular Oncol.*, vol. 43, pp. 195–209, 2020.

[8] K. Sugawara et al., "The impact of pretreatment esophageal stenosis on survival of esophageal cancer patients," *Ann. Surg. Oncol.*, vol. 30, no. 5, pp. 2703–2712, 2023.

[9] Q. Zhao, J. Yu, and X. Meng, "A good start of immunotherapy in esophageal cancer," *Cancer Med.*, vol. 8, no. 10, pp. 4519–4526, 2019.

[10] Y. Ding, X. Lei, B. Liao, and F.-X. Wu, "Predicting miRNA-disease associations based on multi-view variational graph auto-encoder with matrix factorization," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 1, pp. 446–457, Jan. 2022.

[11] S. Sihag et al., "Safety and feasibility of esophagectomy following combined immunotherapy and chemoradiotherapy for esophageal cancer," *J. Thoracic Cardiovasc. Surg.*, vol. 161, no. 3, pp. 836–843, 2021.

[12] T. Tanaka, J. Nakamura, and H. Noshiro, "Promising immunotherapies for esophageal cancer," *Expert Opin. Biol. Ther.*, vol. 17, no. 6, pp. 723–733, 2017.

[13] T. Kojima and T. Doi, "Immunotherapy for esophageal squamous cell carcinoma," *Curr. Oncol. Rep.*, vol. 19, pp. 1–8, 2017.

[14] J. Andreu-Perez, C. C. Poon, R. D. Merrifield, S. T. Wong, and G.-Z. Yang, "Big data for health," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 4, pp. 1193–1208, Jul. 2015.

[15] R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: Current issues and guidelines," *Int. J. Med. Inform.*, vol. 77, no. 2, pp. 81–97, 2008.

[16] W.-T. Wu et al., "Data mining in clinical Big Data: The frequently used databases, steps, and methodological models," *Mil. Med. Res.*, vol. 8, pp. 1–12, 2021.

[17] Z. S. H. Abad, D. M. Maslove, and J. Lee, "Predicting discharge destination of critically ill patients using machine learning," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 3, pp. 827–837, Mar. 2021.

[18] H. Nemat, H. Khadem, M. R. Eissa, J. Elliott, and M. Benaissa, "Blood glucose level prediction: Advanced deep-ensemble learning approach," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2758–2769, Jun. 2022.

[19] F. Divina, A. Gilson, F. Goméz-Vela, M. G. Torres, and J. F. Torres, "Stacking ensemble learning for short-term electricity consumption forecasting," *Energies*, vol. 11, no. 4, 2018, Art. no. 949.

[20] S. Cui, Y. Yin, D. Wang, Z. Li, and Y. Wang, "A stacking-based ensemble learning method for earthquake casualty prediction," *Appl. Soft Comput.*, vol. 101, 2021, Art. no. 107038.

[21] K. Cao et al., "Development of immune gene pair-based signature predictive of prognosis and immunotherapy in esophageal cancer," *Ann. Transl. Med.*, vol. 9, no. 20, 2021.

[22] A. Petrillo and E. C. Smyth, "Immunotherapy for squamous esophageal cancer: A review," *J. Personalized Med.*, vol. 12, no. 6, 2022, Art. no. 862.

[23] L. Yan et al., "Prognostic value of lactate dehydrogenase in advanced esophageal squamous cell carcinoma patients with immunotherapy," *Chin. Gen. Pract.*, vol. 25, no. 26, pp. 3263–3269, 2022.

[24] Z. Huang et al., "Prognostic significance of platelet-to-albumin ratio in patients with esophageal squamous cell carcinoma receiving definitive radiotherapy," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 3535.

[25] Z. Wang et al., "Efficacy and safety of neoadjuvant immunotherapy in surgically resectable esophageal cancer: A systematic review and meta-analysis," *Int. J. Surg.*, vol. 104, 2022, Art. no. 106767.

[26] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Struct. Biotechnol. J.*, vol. 13, pp. 8–17, 2015.

[27] J. Shen, J. Wu, M. Xu, D. Gan, B. An, and F. Liu, "A hybrid method to predict postoperative survival of lung cancer using improved SMOTE and adaptive SVM," *Comput. Math. Methods Med.*, vol. 2021, no. 1, 2021, Art. no. 2213194.

[28] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *Eur. J. Oper. Res.*, vol. 267, no. 2, pp. 687–699, 2018.

[29] Y. Wang, D. Wang, X. Ye, Y. Wang, Y. Yin, and Y. Jin, "A tree ensemble-based two-stage model for advanced-stage colorectal cancer survival prediction," *Inf. Sci.*, vol. 474, pp. 106–124, 2019.

[30] Y. Feng, X. Wang, and J. Zhang, "A heterogeneous ensemble learning method for neuroblastoma survival prediction," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 4, pp. 1472–1483, Apr. 2022.
[31] P. Kaur, A. Singh, and I. Chana, "BSense: A parallel bayesian hyperparameter optimized stacked ensemble model for breast cancer survival prediction," *J. Comput. Sci.*, vol. 60, 2022, Art. no. 101570.
[32] F. Azam et al., "Performance status assessment by using ECOG (eastern cooperative oncology group) score for cancer patients by oncology healthcare professionals," *Case Rep. Oncol.*, vol. 12, no. 3, pp. 728–736, 2020.
[33] B. C. Leibovich et al., "Scoring algorithm to predict survival after nephrectomy and immunotherapy in patients with metastatic renal cell carcinoma: A stratification tool for prospective clinical trials," *Cancer*, vol. 98, no. 12, pp. 2566–2575, 2003.
[34] C. Pilard, M. Ancion, P. Delvenne, G. Jerusalem, P. Hubert, and M. Herfs, "Cancer immunotherapy: It's time to better predict patients' response," *Brit. J. Cancer*, vol. 125, no. 7, pp. 927–938, 2021.
[35] Y. Dai, W. Qiang, K. Lin, Y. Gui, X. Lan, and D. Wang, "An immune-related gene signature for predicting survival and immunotherapy efficacy in hepatocellular carcinoma," *Cancer Immunol., Immunotherapy*, vol. 70, pp. 967–979, 2021.
[36] D. Yuan and C. D. Elvidge, "Comparison of relative radiometric normalization techniques," *ISPRS J. Photogrammetry Remote Sens.*, vol. 51, no. 3, pp. 117–126, 1996.
[37] I. Cohen et al., "Pearson correlation coefficient," *Noise Reduction Speech Process.*, pp. 1–4, 2009.
[38] R. Nevatia and K. R. Babu, "Linear feature extraction and description," *Comput. Graph. Image Process.*, vol. 13, no. 3, pp. 257–269, 1980.
[39] S. V. Naidu, C. Mullapudi, and H. Y. Patil, "Early diabetes detection using combination polynomial features and SelectKBest classifier," *SPAST Abstr.*, vol. 1, no. 1, 2021.
[40] J. F. Feng, Y. Huang, W. S. Lu, and Q. X. Chen, "Preoperative platelet count in esophageal squamous cell carcinoma: Is it a prognostic factor?," *Langenbeck's Arch. Surg.*, vol. 398, pp. 1115–1122, 2013.
[41] Z. Gao et al., "Efficacy and safety of immunochemotherapy, immunotherapy, chemotherapy, and targeted therapy as first-line treatment for advanced and metastatic esophageal cancer: A systematic review and network meta-analysis," *Lancet Regional Health–Western Pacific*, vol. 38, 2023.
[42] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, pp. 197–227, 2016.
[43] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, pp. 3–42, 2006.
[44] D.-C. Feng et al., "Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach," *Construction Building Mater.*, vol. 230, 2020, Art. no. 117000.
[45] G. Ke et al., "Lightgbm: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30.
[46] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul.-Aug. 1998.
[47] B. H. Menze et al., "A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC Bioinf.*, vol. 10, pp. 1–16, 2009.
[48] M. H. D. M. Ribeiro, R. G. d. Silva, S. R. Moreno, V. C. Mariani, and L. d. S. Coelho, "Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting," *Int. J. Elect. Power Energy Syst.*, vol. 136, 2022, Art. no. 107712.
[49] X. Yin, Q. Liu, Y. Pan, X. Huang, J. Wu, and X. Wang, "Strength of stacking technique of ensemble learning in rockburst prediction with imbalanced data: Comparison of eight single and ensemble models," *Nat. Resour. Res.*, vol. 30, pp. 1795–1815, 2021.
[50] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A comprehensive evaluation of ensemble learning for stock-market prediction," *J. Big Data*, vol. 7, no. 1, pp. 1–40, 2020.
[51] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. 22nd ACM sigkdd Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.
[52] O. Kramer and O. Kramer, "Scikit-learn," *Mach. Learn. Evol. Strategies*, pp. 45–53, 2016.
[53] J. Sun, Y. Yang, Y. Wang, L. Wang, X. Song, and X. Zhao, "Survival risk prediction of esophageal cancer based on self-organizing maps clustering and support vector machine ensembles," *IEEE Access*, vol. 8, pp. 131449–131460, 2020.
[54] B. Ma, G. Yan, B. Chai, and X. Hou, "XGBLC: An improved survival prediction model based on XGBoost," *Bioinformatics*, vol. 38, no. 2, pp. 410–418, 2022.