STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–21.
Published online in International Academic Press (www.IAPress.org)

Multi-modal stacked ensemble model for breast cancer prognosis prediction

Aminu Maigari [1], Zurinahni Zainol [2], Chew Xinying [2,*]

[1 2] *School of Computer Science, University Science Malaysia (USM), Penang, Malaysia*
[2] *School of Computing and Informatics, Albukhary International University, Alor Setar Kedah, Malaysia*

Abstract Breast cancer (BC) is a global health challenge that affects millions of women worldwide and leads to significant mortality. Recent advancements in next-generation sequencing technology have enabled comprehensive diagnosis and prognosis determination using multiple data modalities. Deep learning methods have shown promise in utilizing these multimodal data sources, outperforming single-modal models. However, integrating these heterogeneous data sources poses significant challenges in clinical decision-making. This study proposes an optimized multimodal CNN for a stacked ensemble model (OMCNNSE) for breast cancer prognosis. Our novel method involves the integration of the Tug of War (TWO) algorithm to optimize the hyperparameters of a convolutional neural network (CNN), enhancing feature extraction from three distinct multimodal datasets: clinical profile data, copy number alteration (CNA), and gene expression data. Specifically, we employ the TWO algorithm to optimize separate CNN models for each dataset, identifying optimal values for the hyperparameters. We then trained the three baseline CNN models using the optimized values through 10-fold cross-validation on the METABRIC breast cancer dataset. Finally, we utilize an ensemble learning approach to integrate the models' predictions and apply an SVM classifier for the final prediction. To validate the proposed method, we used the TCGA dataset as an independent test set. Our results demonstrated the effectiveness of the OMCNNSE approach for predicting breast cancer prognosis, achieving high AUC, accuracy, sensitivity, precision, and MCC, and outperforming traditional single-modal models and other state-of-the-art methods.

Keywords Breast cancer prognosis prediction, optimized CNN, Tug of War algorithm, stacked-ensemble learning

DOI: 10.19139/soic-2310-5070-2100

1. Introduction

Hyperparameter optimization in deep learning, particularly in convolutional neural networks (CNNs), is crucial in bioinformatics and other fields. It improves model regularization and performance in cancer data. By tuning hyperparameters such as learning rates, batch sizes, and network architectures, researchers can enhance the ability of CNN models to discern intricate patterns within diverse data sources, including mammography images, genomic data, and clinical information. Integrating diverse data types appeared to be the primary challenge, as it is necessary to carefully identify how each modality contributes to model performance [1]. This complicates the hyperparameter tuning process, leading to an exponential increase in hyperparameters that must be optimized simultaneously. Moreover, the risk of overfitting is heightened in multimodal settings, where models may learn noise from the data rather than generalizable patterns [2]. Overfitting can significantly impact the performance of deep learning. Models lead to inaccurate predictions and reduced model interpretability, which is a critical factor in clinical applications such as breast cancer prognosis [3]. These challenges necessitate the development of an effective deep-learning model for breast cancer prognosis, underscoring the importance of precision and interpretability

*Correspondence to: Dr. Chew Xinying (Email: xinying@usm.my). School of Computer Science, University Science Malaysia (USM), Penang, Malaysia.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

in clinical applications [4] By leveraging advanced optimization techniques and validation strategies to enhance the performance of deep-learning models for multimodal breast cancer prognosis and contribute to improved patient outcomes and personalized treatment strategies [5], offering a glimpse of hope against formidable and life-threatening diseases.

In 2020, there was a global incidence of 2.3 million breast cancer cases among women, with a corresponding mortality rate of 685,000 individuals. Breast cancer is the most detected cancer worldwide, with approximately 7.8 million women being diagnosed within the past five years, making it the world's most prevalent cancer, according to the World Health Organization Report 2021. Breast cancer occurs worldwide in women of all ages after puberty, but the incidence of this disease is increasing later in life. Every 14 seconds around the globe, a woman is diagnosed with breast cancer. Breast cancer is the most common cancer in women and is diagnosed in 140 out of 184 countries worldwide, according to a report from the Breast Cancer Foundation Research (BRCF). Breast cancer mortality did not change much from the 1930s to the 1970s. Only in 1980 did the world realize the improvement in survival rates in the countries that established early detection programs, which gave rise to various treatment alternatives for invasive diseases. Given the extreme aggressiveness of breast cancer, prognostic prediction can be improved by incorporating more than one data modality, such as genetics and clinical profile data. However, the heterogeneity of these data has led to complexity because of the significant differences in dimensionality and data types. The process of integrating these data types into a single predictive model is complex [6, 7, 8]. A prognosis estimates the probability of risk, such as complications or death, occurring over a given period. The breast cancer prognosis model divides patients into short-term and long-term survivors based on five-year cutoff points [9]. Accurate prognosis estimation helps clinicians use this knowledge to make clinical decisions [10, 11]. Short-term survivors may consider aggressive treatments to establish eligibility for care programs. Moreover, a favorable prognosis can assist in improving precision medication regimens for breast cancer patients [10, 11]. Predicting breast cancer prognosis has been challenging due to the lack of fast and effective methods in recent decades [6, 7, 8, 12]. Advanced optimization techniques, such as metaheuristic algorithms for deep learning models, enhance the accuracy and efficiency of breast cancer prognosis. This process extracts relevant features from multimodal datasets, mitigates historical limitations, and provides a fast, effective approach. Hyperparameter optimization leads to improved patient outcomes and personalized treatment strategies.

2. Related Work

2.1. Unimodal Approach

[13] was the first to use the 70-gene profile to improve breast cancer prognosis and adjuvant therapy selection using univariate and multivariate statistical techniques. To classify lymph node-negative individuals by gene activity, [14] employed a statistical technique on 115 training sets and discovered 76 gene signatures, with 60 and 16 for ER-positive and ER-negative patients, respectively. They tested their extracted genes against 171 lymph node-negative samples, which yielded 93 percent sensitivity and 48 percent specificity. A feature selection algorithm was adopted to reduce the high dimensionality of the genetic data and improve the model predictions. [15] used SVM and recursive feature elimination (SVM-RFE) to select and predict genes from [13] 70-gene signatures. The 50-gene signature outperformed the 70-gene signature in terms of accuracy, sensitivity, and specificity. [11] used undersampling and bagging with a decision tree classifier and 10-fold stratified cross-validation to classify patients based on a five-year survival threshold to solve the imbalance in the SEER dataset, improving the AUC of the model. [16] used ensemble machine learning with seven publicly available gene expression datasets to classify genes or gene clusters that may accurately identify cancer cells. The ensemble methods predicted cancerous cells better than the individual algorithms. However, prediction models based on gene expression data have disadvantages; single sources of information typically need more non-universality, uniqueness, and noisy data. Multimodal learning provides a solution to the single modality challenges by combining and integrating complementary data from multiple sources to make a final decision [6, 7, 8, 12, 17, 18, 19, 20, 21, 22].

2.2. Multimodal Approach

As microarray technology has advanced, additional data sources have become available, allowing the use of multiple cancer medical datasets, called multimodal datasets, to predict cancer prognosis [23]. In the multimodal approach, researchers are not limited to gene expression data, as described in the section above, but find it exciting and challenging to learn low-level features from the multimodal data [8]. Many studies in the literature have utilized gene expression and clinical data to explore the benefits of combining these two datasets, extracting complementary and valuable information for precise prediction using a multimodal approach. However, combining these two datasets is challenging since gene expression data are high-dimensional and clinical profile data are low-dimensional with mixed-type data (quantitative, qualitative, and interval) [24]. To address this challenge, researchers have proposed different approaches, including adopting different feature selection algorithms that extract relevant features from combined data to produce hybrid markers, which were reported to increase prognosis prediction [10, 11, 24]. Deep belief network (DBN) models were found in the literature to have good predictive accuracies based on various experiments [25, 26]. Unsupervised learning was also adopted using clustering of gene expression data and combining those clustered genes with clinical profile data, improving the precision of the results [27, 28, 29].

Recently, a new utility kernel for SVM has been proposed to deal with the complexity and heterogeneity involved in multiple modalities, whose significance varies in clinical outcomes. The model is simple and yet empirically proves its efficacy by achieving the highest value on various performance metrics [30]. To deal with the high dimensionality of the multiple models, an intelligent system was developed to understand the related features to the prognosis and make correct predictions. Principal component analysis (PCA) and variational auto-encoders (VAEs) are utilized for dimensionality reduction techniques and followed by support vector machines (SVM) or random forests to make the final prediction. this method was found to be effective in identifying long-term and short-term patients [31]. [32] proposes a novel attention-based multi-modal network to accurately predict breast cancer prognosis. The network uses two intra-modality self-attentional modules and an inter-modality cross-attentional module to capture modality-specific and cross-modality relations. The adaptive fusion block takes full advantage of both modality-specific and cross-modality relations. The method effectively boosts prognosis prediction performance and compares favorably with state-of-the-art techniques, addressing challenges in multi-modal fusion and high computational costs.

Deep learning is a relatively new methodology with outstanding performance in various domains. Furthermore, deep multimodal learning is a relatively new field; the concept employs deep neural networks to learn features across multiple modalities and has been widely applied. For breast cancer survival prediction, three or more modalities were used. [33] developed a deep learning application of a multimodal approach to breast cancer prediction. They proposed a deep neural network called MDNNMD by integrating three different datasets. Their method architecture trains three different deep neural network models for each type of data (clinical profile data, copy number alteration, and gene expression) and integrates them using fusion techniques. Their methods achieved better performance than single models and other existing approaches. To address the drawbacks of MDNNMD, [34] developed the STACKED RF method using the same METABRIC dataset in a stacked ensemble architecture. They replaced DNN models with CNNs (CNN-Clinical, CNN-CNA, and CNN-Expr). [35] improved upon their work [34]; they called the newly proposed methods the SiGaAtCNN using a sigmoid-gated attention CNN. The attention mechanism is a neural architectural component that allows dynamic highlighting of relevant features of the input data. The proposed method is called the SiGaAtCNN STACKED RF, and its metrics show that it outperforms logistic regression (LR), SVM, RF, and MDNNMD models. In the literature, it is determined that CNN is among the most accurate models for predicting breast cancer prognosis using multimodal stacked ensemble methods [34, 35].

Recent advancements in deep-learning ensemble architectures have significantly improved breast cancer survival models using multiple datasets. For instance, [36] introduced a graph convolutional network combined with a

Choquet fuzzy ensemble, while [21] developed an ensemble model (EBCSP) that integrates convolutional neural networks for clinical modalities, deep neural networks for copy number variations, and long short-term memory for gene expression data. [37] presented a novel ensemble architecture that overcomes traditional ensemble limitations by incorporating deviation and support mechanisms. Additionally, [38] created a hybrid ensemble deep learning model to enhance prediction accuracy. To address data imbalance, [39] proposed a deep multi-modal fusion network (DMMFN), and [20] developed an attention-based multimodal deep learning model for analyzing mammography images. Furthermore, [40] introduced a multimodal data adversarial representation framework (MDAR) using an embedding space and a multi-scale bilinear convolutional neural network (MS-B-CNN). [41] proposed a novel deep learning approach using a Twin Convolutional Neural Network (TwinCNN) framework and a binary optimization method to eliminate non-discriminative features, addressing challenges in multimodal breast cancer data analysis. These innovations contribute significantly to precision medicine for breast cancer patients. However, given the multidimensionality of the data, the prediction can be enhanced by determining the optimal CNN hyperparameter values. The performance of a CNN model heavily depends on its parameters. Because there is no single design for all the issues, solving a specific problem using the CNN model requires discovering the ideal values of the hyperparameters. According to the reviewed literature, most studies concentrated on enhancing the CNN architecture for classification accuracy, with only a few addressing the subject of hyperparameter optimization [19]. The CNN hyperparameter optimization method has demonstrated superior performance in different fields, such as computer vision [42, 43], the Internet of Things (IoT) [44], image recognition [45], and pattern recognition [46]. The methodology also recently improved disease classification in areas such as diabetic myelopathy [47], brain tumors [45], and cardiovascular disease [48]. These works in the literature have demonstrated the application of hyperparameter optimization using CNNs [44, 49, 50].

Choosing the best hyperparameter values for a given problem is an NP-hard problem and one of the most difficult challenges in the CNN model. Experts employ a variety of strategies to address this challenge. Grid search, random search, and Bayesian optimization are the three most prevalent hyperparameter optimization strategies. A grid search is a systematic trial and error method in which a domain expert sets and assesses each parameter combination in the model. However, the technique only applies to a few hyperparameters; otherwise, it is computationally expensive. The random search uses random sampling in the search space to locate accurate hyperparameter values, making random search more effective than grid search. Its main problem is repeating the previous test's hyperparameter [42]. Bayesian optimization is the most widely used method for identifying neural network architectures and optimizing hyperparameters [51]. However, because Bayesian optimization is based on Gaussian processes, it is limited to low-dimensional optimization problems. Thus, automatic hyperparameter optimization for CNN models is critical in breast cancer prognosis and diagnosis [52]. Metaheuristic algorithms have had a considerable impact on hyperparameter optimization in a variety of domains. They were created to address various practical problems and have sparked much interest in classification problems. Researchers have widely demonstrated the importance of metaheuristic algorithms in tackling complex optimization challenges in engineering, communications, industry, and the social sciences because of their excellent performance and straightforward application[49]. The Tug of War (TWO) algorithm is one of several metaheuristic algorithms. TWO is a recent method that was introduced in 2016 by [53]; this method outperforms GA and its variants and performs well on various mathematical and engineering benchmark problems. The authors used multiple test functions to evaluate the TWO's performance, and it performed well in various optimization issues. TWO offers several advantages, such as requiring the fewest configurable parameters, being simple to implement with excellent search capacity, and being adaptable to changing the basic TWO version.

The challenge of optimizing hyperparameters in convolutional neural networks (CNNs) is a complex, NP-hard problem that typically requires significant domain expertise and manual effort. This paper aims to address the challenge of efficiently optimizing CNN hyperparameters to enhance feature extraction capabilities by implementing the Tug of War (TWO) algorithm. The TWO algorithm offers a meta-heuristic approach that can navigate the hyperparameter space more effectively, potentially leading to better model accuracy and reliability in predicting breast cancer outcomes. This paper makes the following key contributions:

- Optimization of CNN Hyperparameters: We emphasize the critical role of hyperparameter optimization in CNN models, particularly in the context of exploring the most relevant features from multimodal breast cancer data.

- Application of TWO Algorithm: We demonstrate the effectiveness of the Tug of War (TWO) algorithm in identifying optimal hyperparameter values, significantly improving CNN model predictions. The enhanced model, combined with stacked ensembles, provides a more accurate classification of breast cancer patients into long-term and short-term survival categories.

- Comprehensive Experimental Validation: Extensive experiments were conducted to evaluate the performance of the optimized multimodal CNN stacked ensemble (OMCNNSE) model using the METABRIC dataset and validated the model using the TCGA dataset; the model was rigorously tested with the AUC as the primary evaluation metric and accuracy, sensitivity, specificity, and MCC as secondary measures, and the Ablation test by employing a reliability model test using a calibration curve and Brier score. Our findings reveal that the OMCNNSE model consistently outperforms other methods, offering superior predictive accuracy across all metrics.

This research contributes to the advancement of deep learning methodologies in medical applications, particularly in improving the precision of breast cancer survival predictions through sophisticated hyperparameter optimization techniques.

3. Motivation

Breast cancer survival prediction has significantly benefited from the application of deep learning methods, which have shown promising results in handling complex multimodal data. Previous works, such as [33], introduced a novel architecture using three deep neural networks (DNNs) to extract relevant genetic features from individual datasets, integrating predictions through score-level fusion. This approach marked a significant advancement in predictive performance. Building on this foundation, [34] proposed the STACKED RF model, integrating three CNNs using a random forest as the final classifier. This model removed manual guessing from score-level fusion coefficients and introduced an attention mechanism, further improving the predictive performance. The SiGaAtCNN model [35], another enhancement, outperformed existing ensemble techniques on unbalanced cancer-related data. These advancements highlighted the potential of stacked-ensemble approaches for complex medical predictions. Motivated by these developments, we aim to enhance predictive accuracy by integrating metaheuristic-based hyperparameter optimization, specifically using the Tug of War Optimization (TWO) algorithm, within CNN architectures. Our goal is to extract more relevant features from multimodal breast cancer data, thereby improving the overall prediction accuracy for breast cancer survival.

4. Background details

This study proposed a hybrid method that hybridized deep learning model CNN and Tug of War (TWO) algorithms to optimize CNN hyperparameters. The details of each model are provided in sections 4.1 and 4.2 below.

4.1. Baseline Convolutional Neural Network

The proposed new method adopted a CNN for prediction by extracting the features of each multimodal dataset (i.e., clinical, CNA, and gene expression data). This CNN model is a baseline model hybridized with the Tug of War Optimization (TWO) algorithm that optimizes hyperparameters and generates optimal values for training prediction models. This specific architecture has two convolutional layers with filters and a kernel with an activation function, which introduces nonlinearity to the layer's output. The kernel initialization is assigned as a standard

initialization method for deep learning models. After the second convolutional layer, a dropout layer is added to prevent model overfitting during training. This regularization technique randomly dropped neurons to assist the network in learning the model's robust features. After the dropout layer, a max pooling layer is applied to reduce the complexities of the model network. A flattened layer converts the multidimensional output of the max pooling layer into a 1D vector, which is subsequently fed into a fully connected layer, which expects a 1D input. After the flattened layer, a fully connected layer with hidden units and an activation function are added. The layer's output is then fed into a single output layer with an activation function, which generates the final binary classification output. To prevent overfitting, the output of this layer is subjected to an L2 activity regularization technique. The detailed architecture of the baseline CNN is shown in Figure. 1.

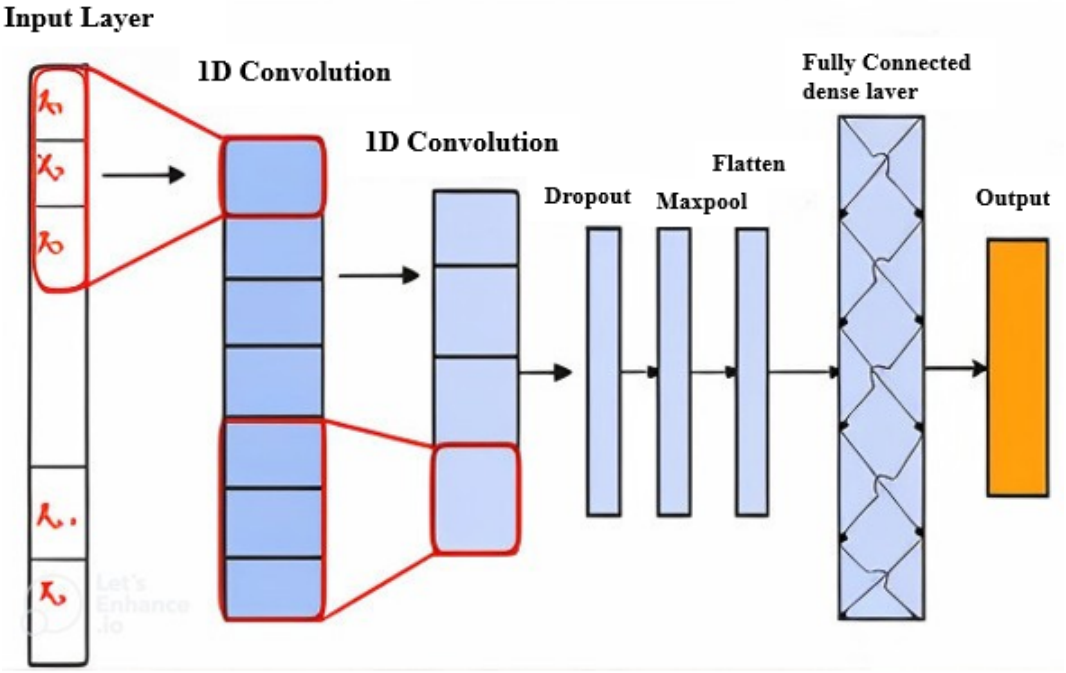

Figure 1. Baseline CNN

The fact that our model is a binary classification problem (for short- or long-term survivors), we adopted binary cross-entropy as a loss function in our CNN baseline model. The loss function is defined as follows:

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)] \tag{1}$$

Where:

- N: Total number of samples
- y_i: Actual label for sample i (either 0 or 1)
- $\hat{y}_i$: Predicted probability (output from the CNN, between 0 and 1)
- $\log$: Natural logarithm

The summation's first term calculates the loss when the true label is 1 (short-term survival), and the second term calculates the loss when the true label is 0 (long-term survival). Our CNN model contains an input layer, two convolution layers, a dropout layer, a max pooling layer, a flattening layer, a fully connected dense layer, and an output layer. Complex CNNs trained on small datasets can overfit. The configurations of the multimodal CNN baseline are the same as those indicated in Table 1 below, except for the input shape, hidden units in the fully

connected dense layer, and optimizers for which the CNA and gene expression multimodal data have different values. We used the AUC metric to evaluate each model.

Table 1. Baseline CNN parameters configuration

Layer type	***Parameters***	***Values***
Input	*Shape*	*(25,1), (200,1), (400,1)*
Conv1D x 2	*filters* *kernel_size* *strides* *padding* *activation* *kernel_initializer* *bias_initializer* *activity_regularizer*	*32* *3* *1* *same* *relu/softplus* *he_uniform* *Constant(0.1)* *l2(0.001)*
Dropout	*rate*	*0.5*
MaxPooling1D	*pool_size* *stride* *padding*	*2* *1* *same*
Flatten		
Dense	*units* *activation* *kernel_initializer* *activity_regularizer*	*43/95/97* *relu* *he_uniform* *l2(0.001)*
Output	*units* *activation* *Loss function* *Optimizer*	*1* *Sigmoid* *binary cross-entropy* *RMSprop/Adam*

4.2. Tug of War (TWO) algorithm

Kaveh and Zolghadr's population-based search approach, Tug of War (TWO), is a metaheuristic algorithm that treats each agent as a team in tug-of-war contests [53]. A team's weight and pulling force on the rope are related to its solution quality. Naturally, the other team must exert at least equal pressure to hold the rope. The convergence operator of the lighter team's TWO algorithm accelerates toward, the heavier team by maintaining a correct exploration and exploitation balance and improving solution quality iteratively. The algorithm steps are as follows;

- **Step 1: Initialization**

The initial position of the team in the search space is assigned at random:

$$x_{ij}^{0} = x_{j,\min} + \text{rand}(x_{j,\max} - x_{j,\min}), \quad j = 1, 2, \ldots, n \tag{2}$$

Where x_{ij}^{0} is the initial value of the i-th and j-th variable, the candidate solutions $x_{(j,\min)}$ and $x_{(j,\max)}$ are the variable's minimum and maximum permissible values, respectively. The random numbers are represented by a uniform distribution within the interval $[0, 1]$.

- **Step 2: Evaluation of candidate designs and weight assignment**

Each candidate solution is handled as a team with the assigned weight and its objective function values assessed, sorted, and saved in a league memory:

$$W_i = \left(\frac{fit(i) - fit_{\text{worst}}}{fit_{\text{best}} - fit_{\text{worst}}} \right) + 1, \quad i = 1, 2, \ldots, N \tag{3}$$

$fit(i)$ is the fitness value of the i^{th} particle, and the i^{th} team penalizes the objective function value for constrained problems. fit_{best} and fit_{worst} are the fitness values for the best and worst candidate solutions, respectively, in the current iteration. The team's weights range from 1 to 2 in the equation.

- **Step 3: Competition and displacement**

Each team competes against all the others one at a time in TWO's algorithm to progress to its new position. A team's pulling force is considered equal to its static friction force $(W\mu_s)$ As a result, the pulling force between teams i and j $(F_{p,ij})$ can be calculated as $\max(W_i\mu_s, W_j\mu_s)$. This definition maintains the heavier team's position. In the iteration, the resultant force on the team as a result of its interaction with a heavier team can be estimated as follows: $F^k_{p,ij}$.

$$F^k_{r,ij} = F^k_{p,ij} - W^k_i \mu_k \tag{4}$$

$F^k_{p,ij}$ is the pulling force between teams i and j in the kth iteration, and μ_k is the coefficient of kinematic friction. As a result, team i accelerates towards team j.

$$a^k_{ij} = \frac{F^k_{r,ij}}{W^k_i \mu_k} g^k_{ij} \tag{5}$$

Where a^k_{ij} is the acceleration of team i towards team j in the kth iteration and g^k_{ij} is the gravitational acceleration constant defined as

$$g^k_{ij} = X^k_j - X^k_i \tag{6}$$

Where X^k_j and X^k_i are the position vectors for candidate solutions j and i in the k-th iteration. Finally, the displacement of team i after competing with team j can be derived as

$$\Delta X^k_{ij} = \frac{1}{2} a^k_{ij} \Delta t^2 + \alpha^k \beta (X_{\max} - X_{\min}) \odot \text{rand}_n(1, n) \tag{7}$$

The equation introduces randomness into the algorithm. A random portion of the search space traveled by team i is interpreted before it stops after removing the applied force. α^k decreases the random position of the team's movement gradually. α is the constant for most applications, ranging from 0.9 to 0.99; larger values decrease the algorithm's convergence speed and help the candidate solutions explore the search space more thoroughly. β is a scaling factor with an interval of $[0, 1]$, and it controls the steps of the candidate solutions when moving in the search space. This parameter should be set lower to search the search space more precisely with fewer steps. For numerical applications, these parameters should be between 0.01 and 0.05; $\mathbf{X}_{\text{max}}$ and $\mathbf{X}_{\text{min}}$ are vectors holding the upper and lower bounds of the design variables, respectively; $\circ$ is element-by-element multiplication; and $\text{rand}_n(1, n)$ is a vector of standard normally distributed random numbers. It should be noted that when team i is lighter than team j, the corresponding displacement of the team j will be equal to zero (i.e., ΔX^k_{ij}). Finally, the total displacement of the team i in iteration k is equal to (with $i \neq j$).

$$\Delta X^k_i = \sum_{j=1}^{N} \Delta X^k_{ij} \tag{8}$$

The team's new position after the iteration is determined as

$$X^{(k+1)}_i = X^k_i + \Delta X^k_i \tag{9}$$

- **Step 4: Updating the League**

The League should be updated after each round of play. The new candidate solutions (team positions) are compared to those of the league's current teams. If the new candidate solution i has a higher objective function value than does the league team, replace it with a new solution.

- **Step 5: Handling the side constraints**

The possibility of lighter team solutions exiting the search space is critical, especially when the values of ΔX are typically larger. Candidate solutions might be returned to their initial feasible location (flyback approach) or randomly generated to solve the problem. The new value of the jth optimization variable of the ith team that violated side constraints in the iteration is defined as

$$x_{ij}^{k} = GB_j + \left(\frac{\text{rand}_n}{k}\right)(GB_j - x_{ij}^{(k-1)}) \tag{10}$$

where GB_j denotes the jth variable of the global best solution (i.e., the best solution thus far); rand_n denotes a random integer generated from a standard normal distribution. There is a minimal chance that the newly developed variable will still be outside the search space. A flyback method is utilized in such instances.

- **Step 5: Termination**

Steps 2 through 5 are repeated until the termination criterion is satisfied

5. Methodology

The proposed methodology for implementing the optimized multimodal CNN for the stacked ensemble model (OMCNNSE) for breast cancer prognosis is described in detail in this section. The section is divided into problem formulation 5.1 and OMCNNSE architecture 5.2, as shown in Fig. 2. The framework comprises data preprocessing, hyperparameter optimization, and training and classification phases.

5.1. Problem formulation

Let the CNN hyperparameters be represented by a vector h as follows:

$$h = (h_1, h_2, h_3, h_4, h_5, h_6, h_7) \tag{11}$$

Where $h_1 - h_7$ represents the hyperparameters to optimize, such as batch size, number of epochs, etc. The vector h is encoded as a solution S where each element corresponds to an encoded value of the hyperparameters:

$$S = [S_1, S_2, S_3, S_4, S_5, S_6, S_7] \tag{12}$$

The decoding function f_{decode} maps the solution vector S to the hyperparameter vector h as follows:

$$h = f_{\text{decode}}(S) \tag{13}$$

- Objective function

The objective function $Acc(h)$ is defined as the classification accuracy of the CNN model with hyperparameters h on the validation dataset. The goal of the optimization is to maximize the accuracy:

$$\text{maximize } Acc(h) \tag{14}$$

Subject to:

$$h \in H \tag{15}$$

Where H is the feasible hyperparameter space defined by the valid ranges for each hyperparameter.

- Proposed CNN-TWO Algorithm Formulation

The proposed hybrid CNN-TWO algorithm can be formally described as:

$$h^* = \arg\max_{s \in S} \text{Acc}(f_{\text{decode}}(s)) \tag{16}$$

Subject to:

$$s_i(t+1) = s_i(t) + \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{W_j + W_i}{W_j - W_i} \cdot (s_j - s_i) \tag{17}$$

where h^* represents the optimal hyperparameter configuration, and S is the space of all possible binary-encoded solution vectors.

This mathematical formulation succinctly captures the encoding of the problem, the objective function, and the application of the TWO algorithm to solve the hyperparameter optimization problem in CNN models. Algorithm 1 describes the pseudocode that demonstrates the CNN-TWO hybrid algorithm that combines Convolutional Neural Networks (CNNs) with the Tug of War Optimization (TWO) algorithm for efficient hyperparameter tuning. CNNs are used for breast cancer prediction, while the TWO algorithm optimizes hyperparameters by treating each configuration as a binary string representing a team in a tug-of-war game. The fitness of each team is based on the CNN's classification accuracy, with teams influencing each other's configurations until the optimal set of hyperparameters is found. This approach enhances CNN performance by leveraging the TWO algorithm's ability to effectively explore the hyperparameter space.

5.2. OMCNNSE framework

We developed the proposed method using a framework with three phases, as shown in Fig. 2. This method consists of data preparation, hyperparameter optimization, and training and classification steps, which are explained in detail below.

1. Data preprocessing stage – The proposed methodology utilizes the METABRIC trial dataset, which consists of clinical data, copy number alterations, and gene expression data. Preprocessing techniques and feature selection were conducted, and the details are presented in the experimentation section below. The data are split into three groups in this phase—training, validation, and testing—with ratios of 60 Percent, 20 Percent, and 20 Percent, respectively. We trained the baseline CNN model with training splits and evaluated it with a validation split. The testing split was saved for future validation of the proposed model during the training and classification stage.

2. Hyperparameter Optimization Stage – In this section, the Tug of War (TWO) algorithm was used to optimize the baseline CNN model hyperparameters to predict breast cancer prognosis. The proposed model optimized seven (7) CNN hyperparameters, including activation functions, batch size, epoch, optimizer, learning rate, network weight initial, and number of hidden units in the fully connected dense layer. Table 2 details the seven CNN hyperparameters and TWO parameters used in the proposed methodology and their value ranges. To obtain the optimal hyperparameter values, we greedily trained each CNN model corresponding to each subset of the data using two-dimensional (TWO) algorithms with preprocessed data. The TWO algorithm uses accuracy metrics as an objective function and has the constraint of maximizing the accuracy metric while minimizing the loss function. After several trials, we found that epoch 5 and pop-size 20 produced the best hyperparameter values for each CNN baseline model (i.e., CLN-CNN-TWO, CNA-CNN-TWO, and GEXP-CNN-TWO); therefore, we set these parameter values as the stopping criteria for the model. Table 3 describes the specific parameter configuration details of the OMCNNSE. The values in Table 3 below are particular to the models developed after applying the

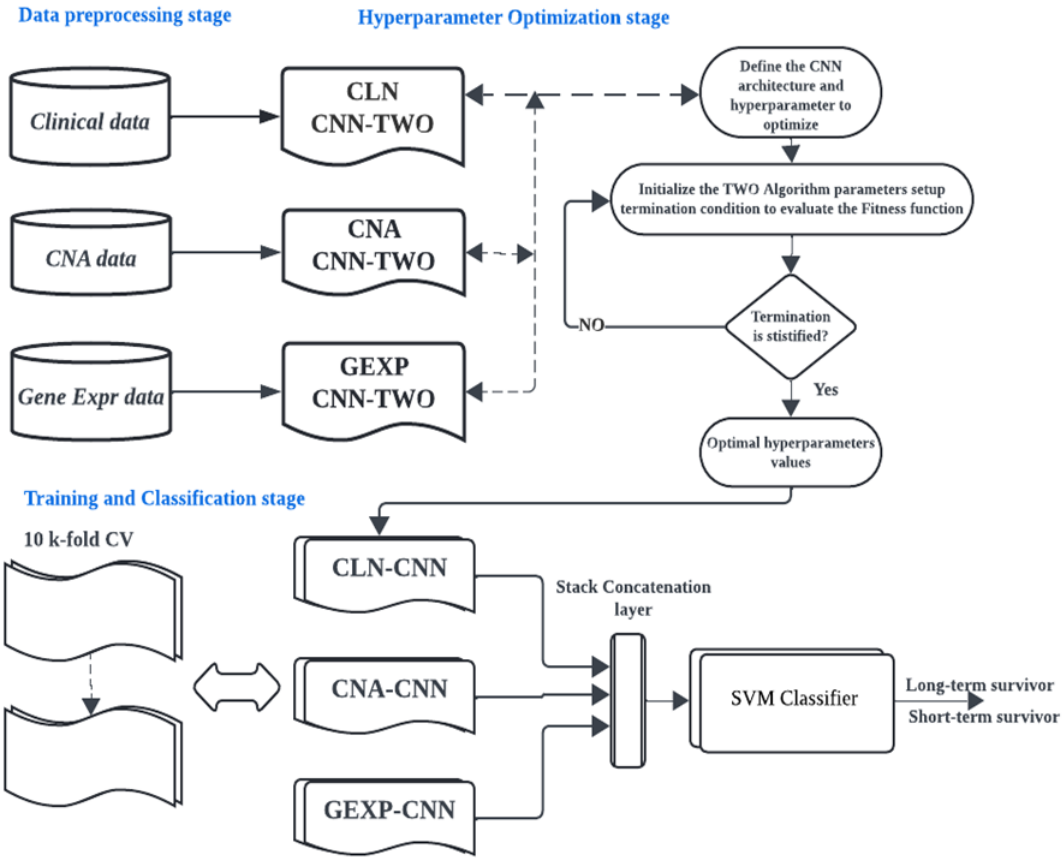

Figure 2. OMCNNSE Framework

TWO algorithm to optimize the CNN baseline models that improve breast cancer prognostic prediction.

3. Training and classification stage for OMCNNSE - For the proposed method, OMCNNSE is used to generalize well on new unseen data, and we apply 10 K-fold cross-validation, which prevents overfitting and optimizes model performance in deep learning models. The stratified K-fold split the dataset (CLN, CNA, and GEXP) into ten equal parts while maintaining the same proportion of classes in each fold. The data are divided into training and testing sets for each fold using the indices obtained from the k-fold split. The model was trained on the training set and evaluated on the testing set. This process is repeated for all ten folds, and the mean accuracy across all folds is calculated. For the clinical, CNA, and gene expression datasets, we designed three different CNN baseline models for all three data modalities: CLN-CNN, CNA-CNN, and GEXP-CNN. We used the AUC metric as the primary performance criterion. We extracted features from the hidden layers of our CNN models. These features are concatenated to form a stacked feature set, which is later used as input to the SVM classifier that generates the final prediction.

Table 2. CNN and TWO default parameter values

#Sn	*Parameters*	*Value ranges*
1.	***CNN Hyperparameters***	
	Batch size	64 - 255.36
	Epoch	50 – 209.9
	Optimizers	['SGD','RMSprop','Adagrad','Adadelta', 'Adam', 'Adamax', 'Nadam']
	Network weight initials	['uniform', 'lecun_uniform', 'normal', 'normal','zero', 'glorot_normal','glorot_uniform','he_normal','he_uniform']
	Activation functions	['softmax', 'softplus', 'softsign', 'relu', 'tanh', 'sigmoid', 'hard_sigmoid', 'linear']
	Learning rate	0.01 – 6.99
	Hidden units	20 – 200
2.	***TWO Algorithm-specific parameters***	
	Epoch	5 – 1000
	Pop_size	20 - 1000

Table 3. Model-optimized parameter values

1. CNN Hyperparameters					
#Sn	***Parameters***	**Values**			
		CLN Model	**CNA Model**	**Gexp Model**	
1.	*Batch size*	*64*	*64*	*64*	
2.	*Epoch*	*70*	*110*	*90*	
3.	*Optimizers*	*RMSprop*	*Adam*	*RMSprop*	
4.	*Network weight initials*	*he_uniform*	*glorot_uniform*	*he_uniform*	
5.	*Activation functions*	*sigmoid*	*sigmoid*	*softplus*	
6.	Learning rates	*0.133*	*0.063*	*0.036*	
7.	Hidden units	***43***	*97*	*95*	
2. TWO algorithm-specific parameters					
	Parameters	***Range of values***			
1.	Epoch	*5*			
2.	Pop_size	*20*			

6. Experiments

This section presents the experimental datasets, followed by the feature selection process, a comparison of different methods, evaluation metrics, and implementation settings.

6.1. Dataset

originally described We used the dataset from GitHub: Https://github.com/USTC-HIlab/MDNNMD repository; the METABRIC dataset was preprocessed before it contained data from 1980 original METABRIC trials of breast cancer patients [54]. Clinical, CNA, and gene expression profile data from breast cancer patients were obtained. A total of 1489 and 491 patients were long-term survivors (0) and short-term survivors (1), respectively. The survival classes were derived using a five-year threshold; the dataset included a median age of 61 years for diagnosed patients and an average survival of 125.1 months. CNA and gene expression data had unknown and null values during preprocessing—the weighted nearest neighbor algorithm [55] was used to remove and normalize these unwanted data points [5]. The gene expression features were further discretized and classified as underexpressed

Algorithm 1 CNN-TWO Hybrid Algorithm

Initialize:
Generate initial population of N binary strings: $S = \{s_1, s_2, \ldots, s_N\}$
Initialize weights W_i for each s_i based on initial fitness (Acc)
Evaluate Fitness:
for each s_i in S **do**
    Decode s_i to hyperparameters $h_i = \text{decode}(s_i)$
    Calculate fitness $F(s_i) = \text{Acc}(h_i)$
end for
TWO Optimization Loop:
while termination condition not met **do**
    Sort S by $F(s_i)$ and update weights W_i
    for each s_i in S **do**
        for each $s_j \neq s_i$ in S **do**
            Calculate force F_{ij} between s_i and s_j
            Update position s_i using: $s_i = s_i + \sum F_{ij}$
        end for
        Apply side constraint handling to s_i
    end for
    Recalculate fitness $F(s_i) = \text{Acc}(\text{decode}(s_i))$
end while
Selection:
Return the $h^\star$ corresponding to $\max(F(s_i))$ as optimal hyperparameters

Figure 3. Algorithm 1

(-1), baseline (0), or overexpressed (1). With five discrete values, the CNA features remain unchanged (2, 1, 0, 1, 2). The authors used min-max normalization [10] in the range of [0, 1] for clinical data. Table 1 contains the general information on the METABRIC breast cancer dataset.

Table 4. Metabric dataset summary

Cut-off (Years)	***5***
Total # of patients	*1980*
Lon-time survivors	*1489*
Short-time survivors	*491*
The median age in diagnosis	*61*
Average survival (months)	*125.1*

6.2. Feature Selection

Here are the details of the feature selection procedures originally described by [33], who performed feature selection on CNA and gene expression data due to the prominent feature length. CNA and gene expression were detected in 26000 and 24000 samples, respectively. The mRMR [56] reduces the dimensions in the gene expression and CNA datasets to avoid the curse of dimensionality. The area under the curve (AUC) values of various feature subsets are observed, and the best feature subset is chosen. For feature selection, a gradational strategy of 100 is

used. The mRMR algorithm chooses the best features from 100 – 500 to determine the highest AUC values. The CNA profile included 200 genes, while the reduced gene expression group included 400 selected genes, which had the highest AUC. The final clinical data included 25 clinical features [35]. Table 5 presents detailed information on the features selected.

Table 5. Selected features

Data Category	***Total Features***	***Selected Features***
Clinical	*27*	*25*
Gene Expression	*24368*	*400*
CNA	*26298*	*200*

6.3. Comparison

We compared OMCNNSE to six robust state-of-the-art models to validate our proposed model. The following are the descriptions of the benchmark work:

- Stacked-RF: This method proposed a stacked ensemble framework using three CNN models; in stage one, they used a CNN to extract features from multimodal data (i.e., clinical, CNA, and gene expression data). Stage two uses the extracted features as input to the stack-based ensemble model for final predictions via the random forest (RF) algorithm [34].
- SiGaAtCNN: This method updated the stacked-RF model and introduced a sigmoid-gated attention convolutional neural network (SGACNN) instead of a CNN to predict the prognosis of breast cancer patients [35].
- MDAR: This study considers modality-invariant embedding space to effectively integrate multimodal data to reduce the modality gap; the proposed model is called multimodal data adversarial representation framework (MDAR) [40].
- DMMFN: A deep multi-modal fusion network (DMMFN) that developed a two-layer one-dimensional Convolutional neural network and the bi-directional long short-term memory network using gated multimodal units to obtain fusion features [39].
- GatedAtt: A methodology that employs multimodal data and generates insightful characteristics using a two-phase model; the first phase generates the stacked features using a sigmoid-gated attention convolutional neural network, and the second phase uses flattened, dense, and dropout processes for bi-modal attention. [20].
- MDNNMD: This work inspired the multimodal breast cancer prognosis approach, in which the author used three distinct deep neural networks to train multidimensional datasets (i.e., clinical data, gene expression, and CNA profile) and then used score-level fusion to aggregate the three models' prediction results and generate the final prediction. Their method is unique in terms of architectural design [33].
- Bayesian optimization: A widely used method for identifying neural network architectures and optimizing model hyperparameters [51].

6.4. Evaluation metrics

We evaluated the performance of our method using the receiver operating characteristic (ROC) curve [57], which was plotted between the false positive rate (1-specificity) and the true positive rate (sensitivity) using the decision

threshold. The AUC is the primary performance indicator, whereas sensitivity (Sn) [57], specificity (Sp) [57], accuracy (Acc) [57], precision (Pre) [57], and Matthew's correlation coefficient (MCC) [55] are the secondary performance metrics of the model.

$$Sn = TP/(TP + FN) \tag{18}$$

$$Sp = TN/(TN + FP) \tag{19}$$

$$Pre = TP/(TP + FP) \tag{20}$$

$$Acc = (TP + TN)/(TP + TN + FP + FN) \tag{21}$$

$$Mcc = (TPTN - FPFN)/((TP + FN)(TP + FP)(TN + FN)(TN + FP)) \tag{22}$$

In a confusion matrix, TP, FP, TN, and FN represent true positives, false positives, true negatives, and false negatives, respectively.

6.5. Implementation settings

We experimented with a Google Colab setting using the computational resources available. The MEALPY platform was used to implement the TWO optimization. The detailed experimental setup is provided under the training and classification subheading above.

7. Results and Analysis

This section presents a detailed analysis and comparative studies of various methods for predicting breast cancer patient survival. In this comparative analysis, we employed the METABRIC dataset.

7.1. Unimodal comparison

The important metric used to evaluate the efficiency of the proposed method is the AUC value displayed in the ROC curve. A lower AUC is inferior to a higher AUC. Fig. 4 shows the ROC curves for the unimodal models.

We compared the model performances using ROC curves to establish the effectiveness of our proposed single model method over the CNN and DNN models in breast cancer prognosis prediction, as shown in Fig. 3. CNN-TWO-Clinical, CNN-TWO-CNV, and CNN-TWO-Expr achieved 2 percent, 20 %, and 4 % greater AUC values than CNN-Clinical, CNN-CNA, and CNN-Expr, respectively. Similarly, 1 %, 12 %, and 20 % improvements in the area under the curve (AUC) were reported for the DNN-Clinical, DNN-CNA, and DNN-Expr models, respectively. As a result, the stacked feature set of the final prediction for the optimized stacked-based model (OMCNNSE) is selected from the hidden features of the proposed single models.

Table 6 compares the OMCNNSE SVM, OMCNNSE RF, and individually optimized CNN models, such as the CNN-TWO-clinical, CNN-TWO-CNA, and CNN-TWO-Expr models, to provide a detailed measurement of the model's AUC values and the remaining secondary metrics.

Table 6 above also indicates that the OMCNNSE SVM model outperformed the OMCNNSE RF and the other unimodal models in terms of the area under the curve (AUC) and different evaluation metrics. In this experiment, we tested the random forest (RF) and support vector machine (SVM) models as the final classifiers in the proposed

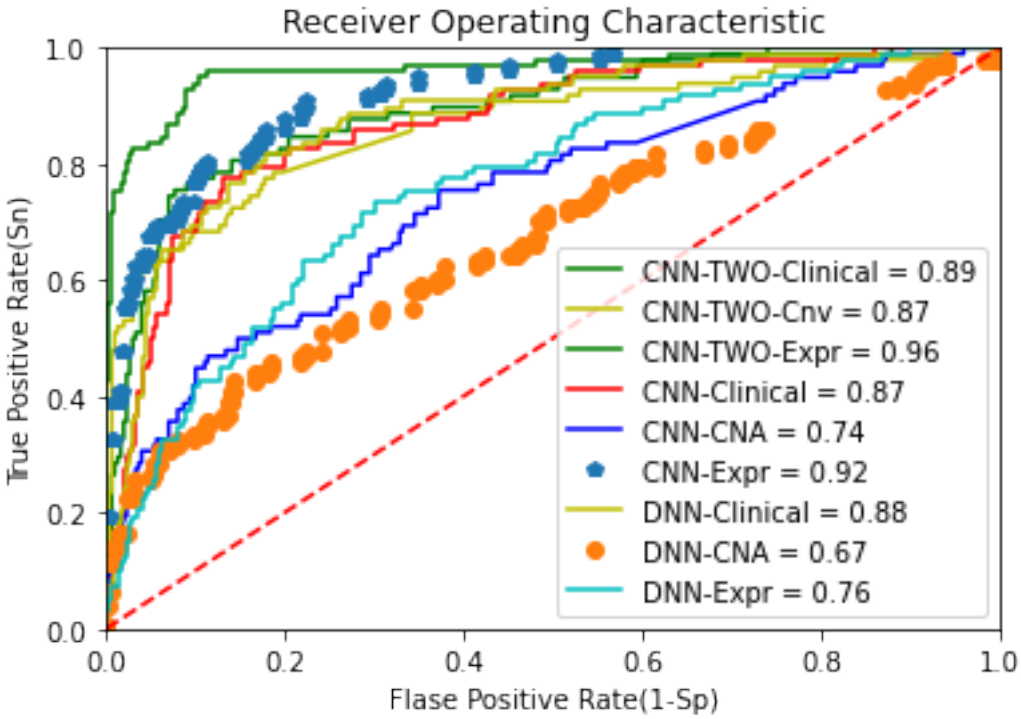

Figure 4. ROC curves and AUC values for unimodal model comparisons

Table 6. OMCNNSE vs. Unimodal result

Models	AUC	Acc	Pre	Sen	MCC
OMMSE SVM	**0.980**	**0.965**	**0.932**	**0.911**	**0.899**
OMMSE RF	0.975	0.939	0.977	0.766	0.829
CLN-CNN-TWO	0.878	0.845	0.782	0.519	0.549
CNA-CNN-TWO	0.871	0.873	0.835	0.607	0.637
GEXP-CNN-TWO	0.951	0.929	0.843	0.878	0.813

ensemble model. OMCNNSE SVM has an AUC of 0.980, which is a 5 percent improvement compared to that of OMCNNSE RF. Notably, the OMCNNSE SVM has an accuracy of 0.965, precision = 0.932, sensitivity = 0.911, and MCC = 0.899. These metric values are 2.6 %, 14.5 %, and 7 % improvement, respectively, except for precision, for which the OMCNNSE RF has a value of 0.977 % compared to 0.911. However, the sensitivity metric representing the classifier's ability to correctly identify patients with a favorable prognosis (i.e., accurately detecting patients with cancer or patients at risk) was greater with the OMCNNSE SVM than with the OMCNNSE RF.

Similarly, the OMCNNSE model outperformed the unimodal model, as described in Table 6. The performance improvements are 12 % AUC, 15 % Acc, 10.4 % Pre, 39.2 % Sen, and 35 % Mcc in the CLN-CNN-TWO model comparisons. Compared to those of the CNN-TWO-CNA, the improvements are 10.9 % AUC, 9.2 % Acc, 9.7 % Pre, 30.4 % Sen, and 26.2 % Mcc. Finally, for the CNN-TWO-Expr dataset, the improvements were 2.9 % in AUC, 3.6 % in Acc, 8.9 % in Pre, 3.3 % in Sen, and 8.6 % in Mcc.

We also compared our proposed OMCNNSE method with two (2) inspired deep learning methods developed by [34, 35] and recently developed four (4) other state-of-the-art methods for breast cancer prognosis prediction, such as MDAR [40], DMMFN [39], GatedAtt [20], and MDNNMD [33]. Fig. 5 below presents a bar chart for the popular state-of-the-art methods and our proposed model (OMCNNSE SVM).

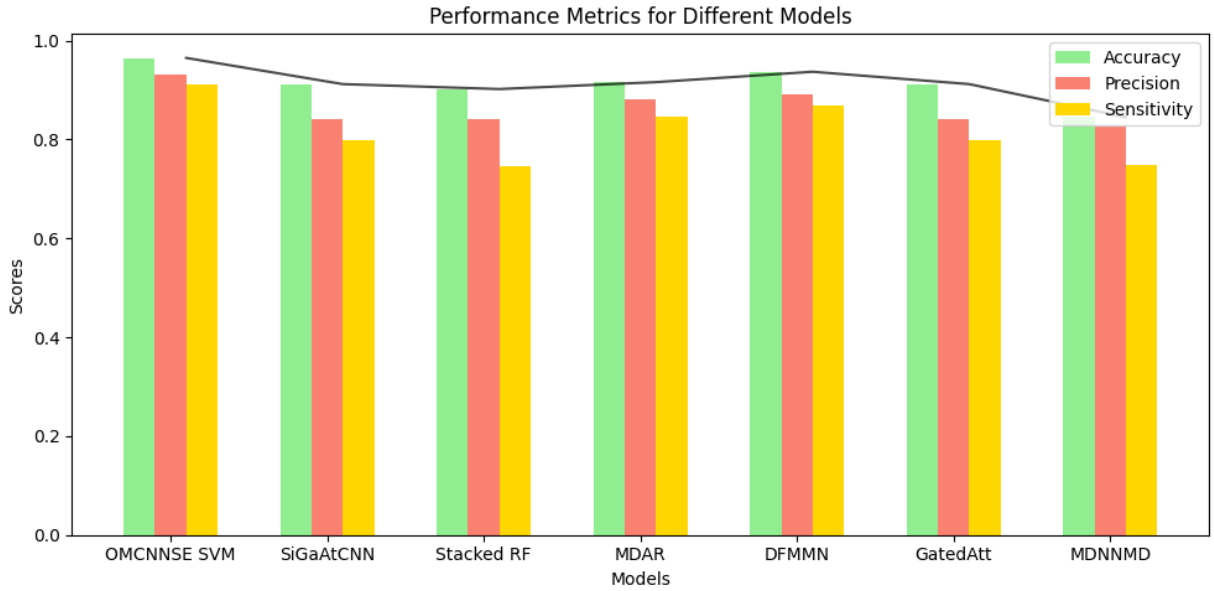

Figure 5. Bar chart of the proposed OMCNNSE SVM and the state-of-the-art models

7.2. Comparison with benchmark prediction methods

Table 7. Performance metrics for different models.

Models	AUC	Acc	Pre	Sen	MCC
OMCNNSE SVM	**0.980**	**0.965**	**0.932**	**0.911**	**0.899**
SiGaAtCNN	0.950	0.912	0.841	0.798	0.762
Stacked RF	0.930	0.902	0.841	0.747	0.730
MDAR	-	0.916	0.882	0.828	0.764
DMMFN	0.964	0.937	0.891	0.845	-
GatedAtt	0.950	0.912	0.841	0.798	-
MDNNMD	0.845	0.826	0.749	0.450	0.486

Table 7 presents details of the comparative analysis of the OMCNNSE SVM model against six other state-of-the-art deep learning models, demonstrating its superior performance across multiple evaluation metrics, including Accuracy, Precision, Sensitivity, Matthews Correlation Coefficient (MCC), and Area Under the Curve (AUC). Starting with the AUC, OMCNNSE SVM achieves the highest value of 0.980, surpassing SiGaAtCNN (0.950) by 3.2%, Stacked RF (0.930) by 5.4%, DMMFN (0.964) by 1.7%, GatedAtt (0.950) by 3.2%, and MDNNMD (0.845) by 16.0%. The AUC value for MDAR is not available, but the comparison still highlights OMCNNSE SVM's superior ability to distinguish between classes. For Accuracy, OMCNNSE SVM scores 0.965, which is 5.8% higher than SiGaAtCNN (0.912), 7.0% higher than Stacked RF (0.902), 5.3% higher than MDAR (0.916),

3.0% higher than DMMFN (0.937), and 5.8% higher than GatedAtt (0.912). This indicates OMCNNSE SVM's higher reliability in correctly classifying samples. In Precision, OMCNNSE SVM achieves 0.932, outperforming SiGaAtCNN (0.841) by 10.8%, Stacked RF (0.841) by 10.8%, MDAR (0.882) by 5.7%, DMMFN (0.891) by 4.6%, and GatedAtt (0.841) by 10.8%. This demonstrates its better capability in reducing false positives. For Sensitivity, OMCNNSE SVM's score of 0.911 is 14.2% higher than SiGaAtCNN (0.798), 22.0% higher than Stacked RF (0.747), 7.8% higher than both MDAR (0.828) and DMMFN (0.845), and 14.2% higher than GatedAtt (0.798). This shows OMCNNSE SVM's effectiveness in correctly identifying true positives. Regarding MCC, OMCNNSE SVM leads with a score of 0.899, surpassing SiGaAtCNN (0.762) by 18.0%, Stacked RF (0.730) by 23.1%, MDAR (0.764) by 17.7%, and MDNNMD (0.486) by 85.1%. The MCC values for DMMFN and GatedAtt are not available, but OMCNNSE SVM's performance in this metric further emphasizes its overall balance between sensitivity and specificity.

Overall, the OMCNNSE SVM model consistently outperforms the other state-of-the-art deep learning models in terms of AUC, Accuracy, Precision, Sensitivity, and MCC, highlighting its robustness and effectiveness in evaluating performance across a broad range of metrics and resulting in more accurate, reliable, and clinically valuable predictions. These improvements enhance decision-making in breast cancer treatment and survival analysis, promoting personalized medicine in oncology and improving patient outcomes.

7.3. Comparison of OMCNNSE with the Bayesian Optimization Algorithm

The Tug of War Optimization (TWO) algorithm outperforms Bayesian Optimization (BO) in model optimization, with the "stackedTWO" model achieving an AUC of $0.950\,\%$, a $4.1\,\%$ higher accuracy than the "StackedBO" model. The TWO model also achieves a $7.6\,\%$ improvement in accuracy, indicating a higher predictive capability. Its precision score is $8.6\,\%$ higher than the "StackedBO" model, indicating a greater ability to correctly identify positive instances. Sensitivity is $29.6\,\%$ higher compared to the "StackedBO" model. The MCC for stackedTWO is $0.899\,\%$, outperforming StackedBO by $26.1\,\%$. These significant percentage differences highlight the TWO algorithm's superior effectiveness in CNN model hyperparameter optimization using breast cancer survival multimodal datasets leading to more accurate, precise, and reliable predictions compared to Bayesian Optimization.

Table 8. Performance metrics for TWO vs BO.

Models	AUC	Acc	Pre	Sen	MCC
stackedTWO	**0.950**	**0.965**	**0.932**	**0.911**	**0.899**
StackedBO	0.913	0.897	0.858	0.703	0.713

7.4. Validation

The TCGA-BRCA dataset [58], which contains 1080 breast cancer patients with CNA, gene expression, and clinical profile data, was used to validate our proposed model. The data are processed and detailed in the Data section above, and the dataset samples include 250 long-term survivors and 830 short-term survivors. We trained and evaluated the OMCNNSE SVM model on the dataset using 10-fold cross-validation; Table 9 contains the details of the results observed for our model and other common approaches. The OMCNNSE SVM model, when compared to other single models using the TCGA dataset, demonstrates superior performance with an AUC of 0.950, accuracy of 91.0%, precision of 84.9%, sensitivity of 73.7%, and an MCC of 73.4%. These results outperform all other models, highlighting their effectiveness in classifying and predicting outcomes. When comparing the OMCNNSE SVM's performance between the METABRIC and TCGA datasets, the model shows a slight decrease across all metrics: AUC drops by 3.1% (from 0.980 to 0.950), accuracy by 5.7% (from 96.5% to 91.0%), precision by 8.9% (from 93.2% to 84.9%), sensitivity by 19.1% (from 91.1% to 73.7%), and MCC by 18.4% (from 89.9% to 73.4%). Despite these decreases, the model remains robust, with strong predictive capabilities on unseen data. The model's high AUC and accuracy across datasets confirm its reliability, though

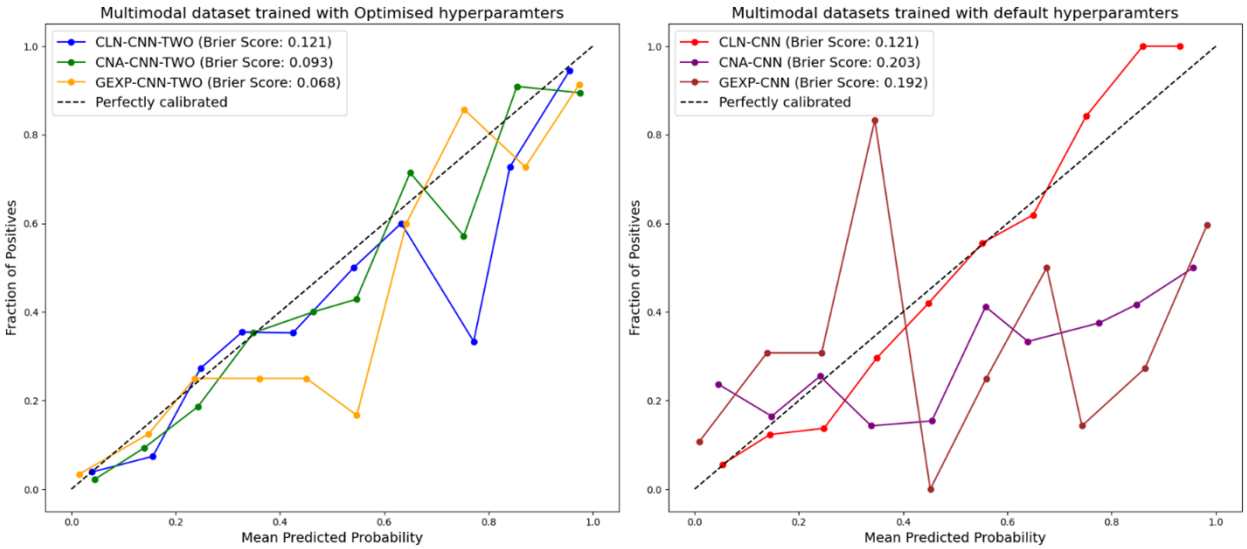

Figure 6. OMCNNSE reliability curve

minor refinements to improve sensitivity and MCC could enhance its generalization to diverse datasets. OMCNNSE

Table 9. Performance metrics for different models.

Models	AUC	Acc	Pre	Sen	MCC
OMCNNSE SVM	**0.950**	**0.910**	**0.849**	**0.737**	**0.734**
CLN-CNN-TWO	0.716	0.781	0.569	0.231	0.257
CNA-CNN-TWO	0.917	0.861	0.844	0.494	0.574
GEXP-CNN-TWO	0.945	0.879	0.948	0.506	0.637

SVM outperformed the other models in all the evaluation metrics of the individual unimodal models and indicated that the proposed model's robustness in classifying independent data was good, with a performance that was close to that of the experimental dataset.

7.5. Model reliability test

This section presents a calibration curve to test the reliability of the OMCNNSE model for breast cancer prognosis and highlights the importance of hyperparameter optimization in deep learning algorithms. The calibration test evaluates how well the model's predicted probabilities of patient survival (e.g., long-term or short-term) align with the outcomes. In addition, a statistical Brier score measures the model's overall accuracy, with lower scores of 0 indicating better calibration and 1 indicating poor accuracy.

Fig. 6 above compares the optimized and default models, revealing that the optimized model demonstrates superior calibration along the diagonal line, indicating accurate alignment between the predicted probabilities and actual outcomes. In contrast, the default model exhibited more variation along diagonal lines, except for the clinical model. The Brier score concurred with the graphical data-generated scores: the CNA-CNN-TWO model had a score of 0.093, whereas the CNN-CNA model had a score of 0.203; the GEXP-CNN-TWO model had a score of 0.068, whereas the CNN-GEXP had a score of 0.192, except for CLN-CNN-TWO and CNN-CLN, which exhibited identical scores of 0.121. This outcome underscores the influence of hyperparameter optimization on genomic data, specifically in the context of copy number alteration (CNA) and gene expression (GEXP).

8. Conclusion

In conclusion, our study highlights the effectiveness of the OMCNNSE model, offering a significant advancement in breast cancer prognosis by optimizing multimodal data integration. The Tug of War (TWO) algorithm's role in feature extraction from clinical, CNA, and gene expression data has proven to be a robust solution for improving predictive accuracy. This multimodal approach addresses the limitations of traditional single-modal models, presenting a comprehensive view of cancer progression that can be directly applied to personalized therapy and patient management. The results underscore the potential of such models to reshape clinical decision-making by integrating diverse data sources, thus offering a holistic view of patient health. To overcome challenges such as small data limitations, future research should expand into transfer learning, data augmentation, and the incorporation of other modalities, including radiomics and histopathological images, further enhancing model generalizability across different patient populations.

However, deep learning interpretability remains a critical hurdle, particularly in clinical domains like cancer prognosis. While deep learning models such as OMCNNSE deliver high performance, their "black-box" nature can hinder clinical trust and adoption. Solutions emerging from the literature, such as attention mechanisms, saliency maps, and model-agnostic techniques, offer pathways to improve interpretability. In multimodal cancer research, these interpretability tools can be adapted to reveal the specific features contributing to survival predictions across various data types, such as gene expression patterns or specific CNA markers. Ensuring that clinicians understand and trust the decision-making process is vital, making interpretability and transparency an ongoing focus. Alongside this, ethical considerations, such as data privacy and consent, must be addressed, especially as cancer datasets continue to grow in complexity and scope. Through these measures, the clinical adoption of deep learning models in cancer care can be more widespread and impactful.

Acknowledgement

This work is funded by the Ministry of Higher Education Malaysia, Fundamental Research Grant Scheme [Grant Number: FRGS/1/2022/STG06/USM/02/4], for the project entitled "Efficient Joint Process Monitoring using a New Robust Variable Sample Size and Sampling Interval Run Sum Scheme".

REFERENCES

1. O. Grigas, "Positive effect of super-resolved structural magnetic resonance imaging for mild cognitive impairment detection," *Brain Sci.*, vol. 14, no. 4, p. 381, 2024.
2. Z. Wang, R. Li, M. Wang, and A. Li, "Gpdbn: Deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction," *Bioinformatics*, vol. 37, no. 18, pp. 2963–2970, Sep 2021.
3. K. Jabeen and et al., "Bc2netrf: Breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-jaya controlled regula falsi-based features selection," *Diagnostics*, vol. 13, no. 7, p. 1238, 2023.
4. U. K. Lilhore, "A precise model for skin cancer diagnosis using hybrid u-net and improved mobilenet-v3 with hyperparameters optimization," *Sci. Rep.*, vol. 14, no. 1, 2024.
5. Y. Sun, "Streamlining ocean dynamics modeling with fourier neural operators: A multiobjective hyperparameter and architecture optimization approach," *Mathematics*, vol. 12, no. 10, p. 1483, 2024.
6. W. Guo, W. Liang, Q. Deng, and X. Zou, "A multimodal affinity fusion network for predicting the survival of breast cancer patients," *Front. Genet.*, vol. 12, Aug 2021.
7. Z. Wang, R. Li, M. Wang, and A. Li, "Gpdbn: Deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction," *Bioinformatics*, vol. 37, no. 18, pp. 2963–2970, 2021.
8. J. Gao, T. Lyu, F. Xiong, J. Wang, W. Ke, and Z. Li, "Predicting the survival of cancer patients with multimodal graph neural network," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 19, no. 2, pp. 699–709, 2022.
9. K. G. M. Moons, P. Royston, Y. Vergouwe, D. E. Grobbee, and D. G. Altman, "Reporting prognosis and prognostic research: What, why, and how?" *BMJ*, vol. 338, no. May, 2009.
10. O. Gevaert, F. D. Smet, D. Timmerman, Y. Moreau, and B. D. Moor, "Predicting the prognosis of breast cancer by integrating clinical and microarray data with bayesian networks," in *Bioinformatics*, Jul 2006.
11. Y. Sun, S. Goodison, J. Li, L. Liu, and W. Farmerie, "Improved breast cancer prognosis through the combination of clinical and genetic markers," *Bioinformatics*, vol. 23, no. 1, pp. 30–37, 2007.
12. Z. He, J. Zhang, X. Yuan, and Y. Zhang, "Integrating somatic mutations for breast cancer survival prediction using machine learning methods," *Front. Genet.*, vol. 11, no. January, pp. 1–12, 2021.

13. V. D. Vijver, "Numb er 25 a gene-expression signature as a predictor of survival in breast cancer," *New England Journal of Medicine*, vol. 347, no. 25, pp. 1999–2009, 2002.
14. Y. Wang and et al., "Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer," *Lancet*, vol. 365, no. 9460, pp. 671–679, 2005.
15. X. Xu, Y. Zhang, L. Zou, M. Wang, and A. Li, "A gene signature for breast cancer prognosis using support vector machine," in *2012 5th International Conference on Biomedical Engineering and Informatics, BMEI 2012*, 2012, pp. 928–931.
16. A. C. Tan and D. Gilbert, "Ensemble machine learning on gene expression data for cancer classification," *Applied Bioinformatics*, vol. 2, no. 3 Suppl, pp. 1–10, 2003.
17. L. Tong, J. Mitchel, K. Chatlin, and M. D. Wang, "Deep learning based feature-level integration of multi-omics data for breast cancer patients survival analysis," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, Sep 2020.
18. A. Dhillon and A. Singh, "Ebrecap: Extreme learning-based model for breast cancer survival prediction," *IET Syst. Biol.*, vol. 14, no. 3, pp. 160–169, Jun 2020.
19. J. Gao, T. Lyu, F. Xiong, J. Wang, W. Ke, and Z. Li, "Mgnn: A multimodal graph neural network for predicting the survival of cancer patients," in *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, Inc, Jul 2020, pp. 1697–1700.
20. S. Kayikci and T. M. Khoshgoftaar, "Breast cancer prediction using gated attentive multimodal deep learning," *Journal of Big Data*, vol. 10, no. 1, 2023.
21. E. Mustafa, E. K. Jadoon, S. K. uz Zaman, M. A. Humayun, and M. Maray, "An ensembled framework for human breast cancer survivability prediction using deep learning," *Diagnostics*, vol. 13, no. 10, pp. 1–13, 2023.
22. N. A. Othman, M. A. Abdel-Fattah, and A. T. Ali, "A hybrid deep learning framework with decision-level fusion for breast cancer survival prediction," *Big Data Cogn. Comput.*, vol. 7, no. 1, p. 50, 2023.
23. T. Baltruˇ, "Multimodal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 1–20, 2019.
24. L. Hedjazi, M. V. L. Lann, T. Kempowsky-Hamon, F. Dalenc, and G. Favre, "Improved breast cancer prognosis based on a hybrid marker selection approach," in *BIOINFORMATICS 2011 - Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms*, 2011, pp. 159–164.
25. M. Khademi and N. S. Nedialkov, "Probabilistic graphical models and deep belief networks for prognosis of breast cancer," in *Proceedings - 2015 IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015*. Institute of Electrical and Electronics Engineers Inc., Mar 2016, pp. 727–732.
26. C. Author, "Prognosis cancer prediction model," *Volume*, vol. 95, no. 20, pp. 5369–5378, 2017.
27. J. Pittman and et al., "Integrated modeling of clinical and gene expression information for personalized prediction of disease outcomes," 2004, available: www.pnas.orgcgidoi10.1073pnas.0401736101.
28. A. H. Chen and C. Yang, "The improvement of breast cancer prognosis accuracy from integrated gene expression and clinical data," *Expert Systems with Applications*, vol. 39, no. 5, pp. 4785–4795, 2012.
29. M. Zhao, Y. Tang, H. Kim, and K. Hasegawa, "Machine learning with k-means dimensional reduction for predicting survival outcomes in patients with breast cancer," *Cancer Informatics*, vol. 17, Nov 2018.
30. N. Arya, A. Mathur, S. Saha, and S. Saha, "Proposal of svm utility kernel for breast cancer survival estimation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 2, pp. 1372–1383, 2023.
31. N. Arya, S. Saha, A. Mathur, and S. Saha, "Improving the robustness and stability of a machine learning model for breast cancer prognosis through the use of multi-modal classifiers," *Scientific Reports*, pp. 1–10, 2023.
32. H. Liu, Y. Shi, A. Li, and M. Wang, "Multi-modal fusion network with intra- and inter-modality attention for prognosis prediction in breast cancer," *Computers in Biology and Medicine*, vol. 168, p. 107796, 2024.
33. D. Sun, M. Wang, and A. Li, "A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 3, pp. 841–850, May 2019.
34. N. Arya and S. Saha, "Multi-modal classification for human breast cancer prognosis prediction: Proposal of deep-learning based stacked ensemble model," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 2, pp. 1032–1041, 2020.
35. ——, "Multi-modal advanced deep learning architectures for breast cancer survival prediction," *Knowledge-Based Systems*, vol. 221, June 2021.
36. S. Palmal, N. Arya, S. Saha, and S. Tripathy, "Breast cancer survival prognosis using the graph convolutional network with choquet fuzzy integral," *Scientific Reports*, vol. 13, no. 1, pp. 1–11, 2023.
37. N. Arya and S. Saha, "Deviation-support based fuzzy ensemble of multi-modal deep learning classifiers for breast cancer prognosis prediction," *Scientific Reports*, pp. 1–10, 2023.
38. E. K. Jadoon, F. G. Khan, S. Shah, A. Khan, and M. Elaffendi, "Deep learning-based multi-modal ensemble classification approach for human breast cancer prognosis," *IEEE Access*, vol. 11, pp. 85 760–85 769, 2023.
39. H. Yuan and H. Xu, "Deep multi-modal fusion network with gated unit for breast cancer survival prediction," *Computer Methods in Biomechanics and Biomedical Engineering*, 2023.
40. X. Du and Y. Zhao, "Multimodal adversarial representation learning for breast cancer prognosis prediction," *Computers in Biology and Medicine*, vol. 157, 2023.
41. O. N. Oyelade, E. A. Irunokhai, and H. Wang, "A twin convolutional neural network with hybrid binary optimizer for multimodal breast cancer digital image classification," *Scientific Reports*, vol. 14, no. 1, pp. 1–23, 2024.
42. N. Bacanin, T. Bezdan, and I. Strumberger, "Optimizing convolutional neural network hyperparameters by enhanced swarm intelligence," *Algorithms*, vol. 13, no. 3, p. 67, 2020.
43. L. A. L. Özkan, "The analysis and optimization of cnn hyperparameters with fuzzy tree model image classification," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 3, 2022.
44. R. O. Ogundokun, S. Misra, M. Douglas, and R. Damaševiˇ, "Medical internet-of-things based breast cancer diagnosis using hyperparameter-optimized neural networks," 2022.

45. M. A. Amou, K. Xia, S. Kamhi, and M. Mouhafid, "A novel mri diagnosis method for brain tumor classification based on cnn and bayesian optimization," pp. 1–21, 2022.
46. H. Respiration, "Hyperparameter optimization method based on harmony search algorithm to improve performance recognition system," pp. 1–19, 2020.
47. G. Atteia and N. A. Samee, "Cnn-hyperparameter optimization for diabetic maculopathy diagnosis in optical coherence tomography and fundus retinography," pp. 1–30, 2022.
48. E. Kıymaç and Y. Kaya, "A novel automated cnn arrhythmia classifier with memory-enhanced artificial hummingbird algorithm," *Expert Systems with Applications*, vol. 213, p. 119162, 2023.
49. E. H. Houssein, "An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators algorithm," *Volume 5*, pp. 18 015–18 033, 2022.
50. P. Kaur, A. Singh, and I. Chana, "Bsense: A parallel bayesian hyperparameter optimized stacked ensemble model for breast cancer survival prediction," *Journal of Computational Science*, vol. 60, no. January, p. 101570, 2022.
51. T. F. Gonzalez, *Handbook of Approximation Algorithms and Metaheuristics*. CRC Press, 2007.
52. K. Balasubramanian, N. P. Ananthamoorthy, and K. Ramya, "An approach to classify white blood cells using convolutional neural network optimized by particle swarm optimization algorithm," *Neural Computing and Applications*, vol. 34, no. 18, pp. 16 089–16 101, 2022.
53. A. Kaveh and A. Zolghadr, "A novel meta-heuristic algorithm: Tug of war," *Volume 6*, no. 4, pp. 469–492, 2016.
54. C. C. et al., "The genomic and transcriptomic architecture of 2,000 breast tumours," *Nature*, 2012.
55. O. T. et al., "Missing value estimation methods for dna microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.
56. Y. Cai, T. Huang, and L. Hu, "Prediction of lysine ubiquitination with mrmr feature selection and analysis," *Amino Acids*, vol. 42, pp. 1387–1395, 2012.
57. A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, 2018.
58. K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "The cancer genome atlas (tcga): An immeasurable source of knowledge," *Współczesna Onkologia*, vol. 1A, pp. A68–A77, 2015.