Forecasting the Yoga Influence on Chronic Venous Insufficiency: Employing Machine Learning Methods

Xiao Du

College of Physical Education, Hubei University of Education, Wuhan 430205, Hubei, China

Abstract—This investigation introduces a groundbreaking approach to unravel the complexities of Chronic Venous Insufficiency (CVI) by leveraging machine learning, notably the Support Vector Classification (SVC), alongside optimization systems like Dwarf Mon-goose Optimization (DMO) and Smell Agent Optimization (SAO). This pioneering strategy not only aims to bolster predictive Precision but also seeks to optimize personalized treatment paradigms for CVI, presenting a compelling avenue for the advancement of healthcare solutions. The study aims to predict the impact of yoga on CVI using a comprehensive dataset, incorporating demographic information, baseline severity indicators, and yoga practice details. Through meticulous feature engineering, machine learning algorithms forecast outcomes such as changes in symptom severity and overall well-being improvements. This predictive model has the potential to transform personalized CVI treatment plans by offering tailored recommendations for specific yoga practices, optimizing therapeutic approaches, and guiding efficient healthcare resource allocation. Ethical considerations, patient preferences, and safety are highlighted for responsible translation into clinical settings. The integration of SVC with optimization systems presents a novel and promising approach, contributing meaningfully to personalized CVI management and providing valuable insights for current and future practices. The results obtained for VCSS-PRE and VCSS-1 unequivocally highlight the outstanding performance of the SVDM model in both prediction and categorization. The model achieved remarkable Accuracy and Precision values, attaining 92.9% and 93.1% for VCSS-PRE and 94.3% and 94.9% for VCSS-1.

Keywords—*Chronic Venous Insufficiency; yoga; classification; machine learning; Support Vector Classification; smell agent optimization; Dwarf Mongoose Optimization*

I. INTRODUCTION

Chronic diseases, defined by the U.S. National Center for Health Statistics as persisting for three months or more, encompass conditions like cardiovascular disease, cancer, arthritis, diabetes, epilepsy, chronic venous disease (CVD), and obesity. These ailments, characterized by their prolonged nature, are not curable through medication. Major contributors to chronic diseases include the use of tobacco, insufficient physical activity, and unhealthy living and eating habits. As examples of Chronic diseases, cardiovascular diseases [1] result from factors such as poor nutrition, lack of physical activity, and tobacco use. Cancer [2], a group of diseases involving abnormal cell growth, has the potential to spread to different body parts. Diabetes [3], characterized by high blood sugar levels, manifests in various types, including Type 1, Type 2, Prediabetes, and Gestational diabetes, each presenting distinct signs and symptoms. As another instance, CVD is a prevalent condition characterized by a spectrum of clinical manifestations, including spider veins, varicose veins, and active venous ulceration. The condition's etiology involves dysfunction in both superficial and deep venous systems [4].

The escalating risk factors associated with CVD present a growing socio-economic and public health challenge. The rising prevalence of obesity and the aging population are anticipated to contribute to an increased burden of CVD over the coming decade, straining available resources for its management. Focusing on the epidemiological, quality of life, and financial aspects of superficial and deep venous disease, with considerations for future projections, the reported prevalence rates for superficial venous disease exhibit significant heterogeneity, with spider veins affecting up to 80% of the population and varicose veins estimated at 30% [5], [6], [7]. Epidemiological studies encounter variability influenced by study population characteristics and modalities, raising concerns about the realistic estimation of disease prevalence. For instance, venous ulcers, impacting 1–2% of the UK population, particularly in older people, pose challenges due to their difficult treatment and recurrent nature [5]. Evidence suggests that CVD is a progressive condition, emphasizing the importance of early prevention.

Moreover, quality of life is substantially impacted by CVD, as indicated by various assessment tools, with depression rates doubling in CVD patients [8], [9]. The financial burden is notable, too, with venous ulcers alone accounting for a significant percentage of the budget expenditure of countries [10]. All these issues necessitate the importance of substantial care within community settings.

The optimal treatment for the human body is not always found in pharmaceutical interventions. Many individuals have experienced adverse effects associated with medication usage, such as antibiotics influencing genetic variability [11]. These side effects encompass hematologic issues, decreased platelet count, drug-induced fevers, rashes, serum sickness, encephalopathy, seizures, blindness, and pulmonary complications, among others [12]. Considering the myriad negative consequences of pharmaceuticals, there has been a significant shift toward emphasizing yoga in medical research. Numerous surveys indicate a rapid increase in the adoption of yoga. Demographic trends reveal that younger individuals, non-Hispanic whites, those with higher incomes, females, college graduates, and individuals with better health status are more inclined to integrate yoga into their lifelong practices [13].

In the past decade, there has been a growing recognition of the significance of Yoga within the medical research community, with a substantial body of literature exploring its applications in various medical contexts, interventions for positive body image [14], [15], including cardiac rehabilitation [16], and the management of mental illnesses [17]. Notably, Yoga is advocated as a therapeutic approach capable of effectively treating numerous diseases without the reliance on pharmaceutical interventions [18]. The practice of Yoga encompasses a range of exercises that not only enhance physical health but also contribute to the purification of the body, mind, and soul [19]. This involves the performance of various asanas, each representing static physical postures [20]. Systems for learning and self-instruction in Yoga have the potential to promote its widespread adoption while ensuring correct practice [21].

Recent technological advancements in machine learning (ML) and Data mining have led to the development of sophisticated methods for processing medical data. A comprehensive review [22] discussed the specifics, challenges, and potential risks associated with ML models in medicine, and several studies have explored diverse applications of machine learning in healthcare [23], [24], [25], [26].

According to Nafee et al. [27], the purpose of the research was to evaluate how well machine learning models identified acutely sick individuals who were at high risk of venous thromboembolism (VTE) when compared to the IMPROVE score. Data from the APEX study, in which 7513 individuals were randomly assigned to receive betrixaban or enoxaparin, were examined by researchers. They used a variety of candidate models and variables to build a reduced model (rML) and a super ML. Every patient's IMPROVE score was determined. The c-statistic values for the ML and rML algorithms were higher (0.69 for ML, 0.68 for rML, and 0.59 for IMPROVE score), indicating that they outperformed the IMPROVE score in predicting VTE. The machine learning models were also preferred by calibration analysis. Compared to patients in the lowest tertile, those in the highest tertile of estimated VTE risk had considerably higher chances of developing VTE. The study's result was that, when it came to predicting VTE in critically sick patients, machine learning algorithms outperformed the IMPROVE score in terms of discrimination and calibration. Ryan et al. [28] focused on the challenges of effectively predicting deep venous thrombosis (DVT) in hospitalized patients, given the limitations of standard scoring systems. The research made use of data from a large university hospital that included 99,237 patients in ICUs or general wards, 2,378 of whom had DVT while they were hospitalized. Gradient-boosted models is a kind of machine learning method, was used to forecast the probability of DVT at 12- and 24-hour intervals prior to initialization. The in-hospital diagnosis of DVT was the main outcome of interest. With AUROCs of 0.83 and 0.85 for DVT risk prediction at 12- and 24-hour periods, respectively, the ML models showed strong performance. A history of malignancy, viral thromboencephalopathy (VTE), and the internal normalized ratio (INR) at 12 and 24 hours before to the beginning of DVT were shown to be significant predictors of DVT risk. The research emphasized the potential therapeutic advantages of

enhanced risk stratification, indicating that it would allow for more focused administration of preventive anticoagulants and lessen the necessity for intrusive testing in difficult patients. This might thus result in an earlier diagnosis and course of therapy, hence reducing the likelihood of problems like pulmonary emboli and other DVT-related sequelae developing. Kumar et al. [29], this study addressed cardiovascular disease (CVD), which includes disorders marked by constricted or clogged veins that may result in strokes, angina, or heart attacks. The purpose of the research was to assess how well machine learning tree classifiers performed in predicting CVD from patient symptoms. The accuracy and AUC ROC scores of a number of machine learning tree classifiers, such as Random Forest, Decision Tree, Logistic Regression, Support Vector Machine (SVM), and K-nearest neighbors (KNN), were investigated. The Random Forest classifier proved to be very successful in the study of Cardiovascular Disease prediction, with a performance time of 1.09 seconds, a ROC AUC score of 0.8675, and a high accuracy rate of 85%. This shows that, in the context of this investigation, the Random Forest classifier had strong predictive skills in diagnosing CVD based on symptomatology.

The articles mentioned above, as is generally accepted, were noticeably devoid of any optimization techniques that could have been utilized to improve precision and reduce complexity in their predictive models. The lack of optimization strategies incorporated in these studies signifies a substantial deficiency in the predictive analytics methodology utilized. Optimization methodologies are crucial in the process of fine-tuning and refining a predictive model, which ultimately increases their accuracy and decreases their computational complexity. Through the process of algorithm optimization, scholars have the ability to methodically improve the overall efficacy of predictive models, thereby guaranteeing a more precise and streamlined depiction of the latent patterns within the data. By enhancing prediction outcomes and contributing to the enhancement of computational efficiency, optimization techniques enable the development of models that are not only more scalable but also more adaptable to diverse datasets. Fundamentally, the incorporation of optimization methodologies is a critical component in enhancing the predictive modeling procedure, thereby promoting more reliable and efficient results in analyses driven by data. Inspired by all existing literature and considering the gap related to the investigation of ML application in effect detection between Yoga and CVD.

This study aims to construct robust machine-learning models for forecasting the impact of Yoga on CVD, harnessing data from credible sources. The chosen methodology involved the application of the Support Vector Classification (SVC) technique. An inventive strategy was implemented by seamlessly incorporating two optimization algorithms, namely Dwarf Mongoose Optimization (DMO) and Smell Agent Optimization (SAO), infusing the predictive modeling process with a nuanced and sophisticated dimension. SVC was selected as the predictive model for assessing the effects of yoga on CVD due to its proficiency in handling complex datasets and non-linear relationships. It excels in classification tasks, making it well-suited for discerning patterns and predicting

outcomes. The model's robustness, coupled with its ability to capture intricate relationships, makes it an effective choice for predicting the impact of yoga on CVD. A comprehensive analysis of the pertinent data, the model, and the optimizers implemented in Section II will be presented in the subsequent sections. A comprehensive analysis of the metrics-driven models and an in-depth explication of the data will be presented. The outcomes obtained from the training and testing stages will be thoroughly examined in Section III. Discussion is given in Section IV and finally, Section V concludes the paper.

II. MATERIALS AND METHODOLOGY

A. Support Vector Classification (SVC)

Support Vector Classification is an algorithm based on the foundational concept of minimizing risk within the context of support vector machines [30]. It involves applying non-linear transformations to the independent variables and projecting them into a high-dimensional space. Within this space, an optimal hyperplane is created to separate the two classes effectively. The main objective of this hyperplane is to minimize classification errors while maximizing margins, representing the overall distance from the hyperplane to the nearest training samples of each class [31].

The main model is subsequently presented in Eq. (1) to Eq. (3) [32].

$$min_{w,b,\in} \frac{\|W\|^2}{2} + C_{svc} \sum_{i=1}^{N} \in_i \qquad (1)$$

$$y_i(w^T.\emptyset(x_i) + b) \geq 1 - \in_i \qquad i = 1,\ldots,N \qquad (2)$$

$$\in_i \geq 0 \qquad i = 1,\ldots,N \qquad (3)$$

The function $\emptyset(x_i)$ denotes a non-linear transformation that takes each observation, characterized by its explanatory variables x_i, and maps it into a higher-dimensional space. C_{svc} represents a regularization parameter, w symbolizes the weight vector associated with the explanatory variables in the newly defined space commonly referred to as the "feature space." b signifies a bias term, and $\in_i$ are slack variables indicating the gap or distance between individual observations (i) and the margin boundary associated with their respective classes.

Identifying the optimal hyperplane, as outlined in Eq. (4), involves maximizing the margin within the high-dimensional space. This process essentially revolves around minimizing the norm of the weight vector while also reducing the number of misclassified instances. In the end, the labels or output variables signify the class to which each sample belongs.

$$D(x_i) = W^T\varphi(x_i) + b \qquad (4)$$

The dimensionality of the problem influences the magnitude of the primal model, whereas the number of samples influences the dual form. Consequently, when the dimensionality is high enough, it becomes more beneficial to deal with the dual model, as indicated in Eq. (5) to Eq. (7).

$$max_a \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i=1}^{N} a_i a_j y_i y_j K(x_i, x_j) \qquad (5)$$

$$\sum_{i=1}^{N} a_i y_i = 0 \qquad (6)$$

$$0 \leq a_i \leq C_{svc} \qquad i = 1,\ldots,N \qquad (7)$$

A Kernel function, represented as $K(x_i, x_j)$, maps each pair of data points to a corresponding location in the feature space. There are various Kernel functions available, such as linear, polynomial, radial basis, sigmoidal, and others. A crucial requirement for these functions is that they must be symmetric, positive, and semi-definite. Previous research in this field has demonstrated that the radial basis Kernel function, defined in Eq. (8), is particularly well-suited for classification tasks [33]. Consequently, a radial basis Kernel function is utilized in the present approach, with 'γ' serving as a hyperparameter indicating the inverse of the range of influence of the data points identified as support vectors [34].

$$K(x_i, x_j) = \emptyset(x_i)^R \emptyset(x_j) = exp(-\gamma\|x_j - x_i\|) \qquad (8)$$

After solving the model to estimate the weights and the bias term, predictions for new samples can be generated using Eq. (9).

$$SVC \quad y_i = \begin{cases} -1 \ if \ w^T\emptyset(x_i) + b \ \leq 0 \\ 1 \ if \ w^T\emptyset(x_i) + b \ > 0 \end{cases} \qquad (9)$$

B. Smell Agent Optimization (SAO)

The significance of the sense of smell in sustaining life on Earth has been profound since the planet's inception. Many living organisms detect harmful substances in their environment through their olfactory receptors [35], [36], [37]. A common practice in the development of Search and Rescue Agents (SAO) involves integrating the human sense of smell [37], [38], [39]. The SAO's structure is based on three modes derived from the olfactory sense. The initial mode entails detecting and evaluating olfactive molecules to decide whether to pursue or ignore the scent. The second mode builds upon the first to track scent particles and locate their source. The third mode prevents the agent from getting trapped and ensures it can maintain its trail.

1) Sniffing mode: Initiating the process involves randomly selecting a location for the diffusion of odor molecules toward the agent, taking into account that olfactory molecules typically propagate in the direction of their target. The mathematical formula, represented by Eq. (10), can be utilized to initialize the scent molecules.

$$x_i^{(t)} = \begin{bmatrix} x_{(1,1)} & x_{(1,2)} & x_{(1,D)} \\ . & . & . \\ x_{(N,1)} & x_{(N,2)} & x_{(N,D)} \end{bmatrix} \qquad (10)$$

Here, D signifies the total count of decision variables, while N represents the overall number of scent molecules present.

Eq. (10) utilizes a location vector that allows the agent to identify its optimal position within the search space. This optimal location can be determined using Eq. (11):

$$x_i^{(t)} = lb_i + r_0 \times (ub_i - lb_i) \qquad (11)$$

r_0 is a randomly generated number ranging from 0 to 1. In relation to the decision variables, ub and lb represent the upper and lower bounds, respectively.

Eq. (12) is employed to assign a primary velocity for diffusion to each scent molecule originating from the odor source.

$$v_i^{(t)} = \begin{bmatrix} v_{(1,1)} & v_{(1,2)} & v_{(1,D)} \\ . & . & . \\ v_{(N,1)} & v_{(N,2)} & v_{(N.D)} \end{bmatrix} \qquad (12)$$

Every single molecule's scent can potentially signify a feasible solution. The position vector determines the potential solutions, $x_i^{(t)} \in R^N$, as illustrated in Eq. (12), along with the molecular velocity, $v_i^{(t)} \in R^N$, as specified in the same equation. The increase in molecular velocity is achieved through Eq. (13):

$$x_i^{t+1} = x_i^{(t)} + v_i^{t+1} \times \Delta t \qquad (13)$$

The optimization process is progressed by the agent simultaneously when the time interval Δt is set to 1. Eq. (14) is used to determine the fragrance molecules' spatial coordinates:

$$x_i^{t+1} = x_i^{(t)} + v_i^{t+1} \qquad (14)$$

Every scent molecule possesses unique diffusion velocities that facilitate its positional updates and evaporation during scent analysis. Eq. (15) is employed to calculate the adjusted velocity of the scent molecules.

$$v_i^{t+1} = v_i^{(t)} + v \qquad (15)$$

The variable governing the velocity update, denoted as v, is determined by utilizing Eq. (16):

$$v = r_1 \times \sqrt{\frac{3KT}{m}} \qquad (16)$$

The smell fixation factor, denoted by the letter "k," serves to normalize the impact of temperature and mass on the kinetic energy of fragrance molecules. The letters "m" and "T" in this instance denote the smell molecules' mass and temperature, respectively.

The evaluation of the fitness of the scent molecule at the adjusted locations is conducted using Eq. (13). Consequently, the sniffing process is completed, allowing for the determination of the exact location of the agent, denoted as x_{agent}^t.

2) Trailing mode: The second operational mode entails simulating the agent's behavior to locate the source of a particular scent. During the search for the scent source, the agent can identify a new location with a higher concentration of scent molecules through olfactory perception. To explore this newly detected location, the agent utilizes Eq. (17):

$$x_i^{t+1} = x_i^{(t)} + r_2 \times olf \times \left(x_{agent}^t - x_i^{(t)}\right) - r_3 \times olf \\ \times \left(x_{worst}^t - x_i^{(t)}\right) \qquad (17)$$

The term r_2 penalizes the impact of olfaction on x_{agent}^t, while r_3 penalizes the effect of olfaction on x_{worst}^t. Both r_2 and r_3 are numerical values ranging from 0 to 1. In the sniffing mode, the agent records x_{agent}^t and the x_{worst}^t. This data is vital for the algorithm to maintain a balance between exploration and exploitation, as depicted in Eq. (17).

3) Random mode: In scenarios where the distance between scent molecules is notably fragmented, their intensity may vary over time. This variation can confuse the agent, leading to the dissipation of the scent and complicating tracking. The agent's difficulty in retaining trail information may result in being trapped in local minima. In such situations, the agent transitions to the random mode, as represented by Eq. (18):

$$x_i^{t+1} = x_i^{(t)} + r_4 \times SL \qquad (18)$$

The term r_4 penalizes the quantity of step length SL, where SL represents the step length.

Algorithm 1. presents the pseudo-code depicting the SAO method:

Algorithm 1 Smell Agent Optimization
Initialize Parameters
Initialize smell molecules' initial position
Assess fitness
Prepare the location of the agent and the worst position of molecu
While ($Itr < Itr_{max}$) do:
for ($i = 1$ to molecules) do:
for ($j = 1$ to position) do:
update molecules' velocity and position (sniffing)
end for
Assess fitness
if (new fitness is better) then:
Update fitness
Update agent and worst molecules
end if
end for
for ($i = 1$ to molecules) do:
for ($j = 1$ to position) do:
update position (trailing)
end for
Assess fitness
end for
if (new fitness is better) then:
grant new fitness
update position
else
for ($i = 1$ to molecules) do:
for ($j = 1$ to position) do:
Implement random mode
end for
end for
end if
end while
return optimum solution.

C. Dwarf Mongoose Optimization (DMO)

The DMO (Dwarf Mongoose Optimization) algorithm is a stochastic metaheuristic method that operates on a population basis. It derives inspiration from the social and foraging behaviors exhibited by the dwarf mongoose, as documented by Helotage [40].

The DMO's problem-solving approach commences by choosing an initial set of potential solutions within the mongoose colony. This involves generating an initial population of candidate solutions and randomly creating them within the predetermined minimum and maximum limits specified for the particular problem at hand. The stochastic generation of solutions ensures adherence to the defined upper and lower bounds of the problem.

$$k = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d-1} & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d-1} & x_{2,d} \\ \vdots & \vdots & x_{1,1} & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,d-1} & x_{n,d} \end{bmatrix} \quad (19)$$

The symbol d denotes the dimensionality of the underlying problem, while n represents the cardinality of the population. The positional attribute of individual elements in a population is denoted as $x_{i,j}$ and determined by applying Eq. (20)

$$x_{i,j} = unifrnd(Var_{Min}, Var_{Max}, Var_{Size}) \quad (20)$$

The term Var_{Size} is associated with the dimensions and ranges of the problem under consideration. The *unifrnd* function serves as a random number generator, producing numbers with a uniform distribution. Var_{Min} and Var_{Max} represent the lower and upper bounds, respectively.

With two different phases—exploration and exploitation—the DMO algorithm adheres to the usual metaheuristic methodology. Known as intensification, every mongoose does a comprehensive search within the defined region during the exploitation phase. On the other hand, the phrase "exploration phase" refers to a more haphazard quest for novel resources, such as food supplies or sleeping mounds. Three crucial social structures—the alpha group, scout group, and babysitters—allow the DMO algorithm to function during these two stages. The coordination of the solution population's actions, which guarantees efficient search space exploration and exploitation, is greatly aided by these structures. This is a list of all the things that need to be done.

1) Alpha group: To designate the alpha female (α) for leading the family unit, Eq. (21) is employed as a selection method.

$$\alpha = \frac{fit_i}{\sum_{i=1}^{n} fit_i} \quad (21)$$

n represents the current number of mongooses that comprise the alpha group, $peep$ refers to an auditory signal that is produced by a dominant or alpha female mongoose. In addition, bs is utilized to represent the number of individuals within the mongoose group who are tasked with the responsibility of caring for and supervising young offspring.

The sleeping mound demonstrates a positive correlation with a plentiful supply of nutritional ingredients, as calculated by Eq. (22):

$$X_{i+1} = X_i + \varphi * peep \quad (22)$$

φ is a numerical value uniformly distributed within the range of [-1, 1].

During each iteration, of the algorithm, the size and quality of the sleeping mound are assessed, as indicated by mathematical Eq. (23):

$$sm_i = \frac{fit_{i+1} - fit_i}{max\{|fit_{i+1}, fit_i|\}} \quad (23)$$

Upon detecting a previously inactive accumulation, a statistical measure is computed using mathematical Eq. (24):

$$\rho = \frac{\sum_{i=1}^{n} sm_i}{n} \quad (24)$$

2) Scout group: After fulfilling the requirements for participation in a babysitter exchange program, the subsequent step involves a scouting stage. In this stage, an assessment is conducted to identify a suitable sleeping location, contingent on the availability of a specific sustenance source. Acknowledging the tendency of mongooses to avoid reusing previously employed sleeping locations, the scouting group is assigned the task of locating a new sleeping mound to facilitate the ongoing advancement of their exploratory endeavors. Within the context of the DMO algorithm, the mongoose demonstrates a distinctive activity pattern marked by foraging and scouting behaviors. This behavior operates on the premise that increasing the distance covered during foraging activities enhances the likelihood of discovering a new sleeping location. Mathematically, this process is represented by the utilization of Eq. (25) to Eq. (27):

$$X_{i+1} = \begin{cases} X_i - CF * phi * rand * [X_i - \vec{M}] & if \ \rho_{i+1} > \rho_i \\ X_i + CF * phi * rand * [X_i - \vec{M}] & else \end{cases} \quad (25)$$

$$CF = \left(1 - \frac{iter}{Max_{iter}}\right)^{\left(2\frac{iter}{Max_{iter}}\right)} \quad (26)$$

$$\vec{M} = \sum_{i=1}^{n} \frac{X_i \times sm_i}{X_i} \quad (27)$$

$\vec{M}$ denotes the force propelling the movement of the mongoose toward a recently formed sleeping mound, and $rand$ signifies a random number that is uniformly distributed within the range of [-1, 1].

3) Babysitters group: While the scouting and foraging team searches for a suitable location for rest and food, the group dedicated to the well-being of the young offspring remains vigilant in monitoring and caring for them. The pool of available candidates for the babysitter exchange diminishes as certain group members opt to postpone their foraging or

scouting activities until they fulfil the requirements for participating in the exchange program. Algorithm 2 provides the pseudo-code for the DMO algorithm.

Algorithm 2 Pseudo-Code of DMO Algorithm
Set the parameters of the algorithm: *Generate* *for iter = 1: max_iter* *Compute the fitness of the mongoose* *Set time counter C* *Determine the alpha* $\alpha = \dfrac{fit_i}{\sum_{i=1}^{n} fit_i}$ *Obtain a candidate for a food position* $X_{i+1} = X_i + \varphi * peep$ *Guess new fitness of* X_{i+1} *Guess sleeping mound* $sm_i = \dfrac{fit_{i+1} - fit_i}{max\{|fit_{i+1}, fit_i|\}}$ *Compute the sleeping mound average value* $\rho = \dfrac{\sum_{i=1}^{n} sm_i}{n}$ *Compute the movement vector* $\vec{M} = \sum_{i=1}^{n} \dfrac{X_i \times sm_i}{X_i}$ *Exchange babysitters if* $C \geq L$ *Set bs position* *compute fitness* $fit_i \leq \alpha$ *Simulate the scout mongoose's next position.* $X_{i+1} = \begin{cases} X_i - CF * phi * rand * [X_i - \vec{M}] \ if \ \rho_{i+1} > \rho_i \\ X_i + CF * phi * rand * [X_i - \vec{M}] \qquad\quad else \end{cases}$ *Modernize the best solution so far.* *end For* *return the best solution* *end*

D. Data Processing

Involving the extraction of valuable information from vast datasets, data mining, also known as database knowledge discovery, utilizes various techniques. The analysis of extensive data collection is a key aspect of this process, revealing hidden patterns and relationships that can significantly impact decision-making. Data mining approaches often incorporate the use of questionnaires or structured datasets presented in the form of reports. This study systematically extracted data from extant literature, comprising a cohort of 100 male subjects, with meticulous attention to both input and output variables. The input variables, influential in determining Chronic Venous Insufficiency (CVI) levels, encompassed diverse facets, including physical attributes (Age, Height, Weight, Body Mass Index (BMI)), Ankle-Brachial Pressure Index (ABPI), Diabetes Blood pressure type A and B (DBPA and DBPB), Pulse Rate (PR), cardiometabolic and vascular health indices (Systolic Blood Pressure type A and B (SBPA and SBPB), Left and Right Calf Circumstances (LE CA-CIR and RI CA-CIR), Mental Chronic Fatigue Syndrome (CFS MEN), Physical Chronic Fatigue Syndrome (CFS PHY), Hyper-homocysteine Mia (HCY), Left and Right Ankle Circumstances (LE AN-CIR and RI AN-CIR), and the Chronic Venous Insufficiency Questionnaire (CVIQ_total)) [41].

Additionally, factors about living conditions and habits were considered, encompassing parameters such as Sleep quality, smoking status, Alcohol intake, Dietary habits, and the duration of sitting and standing hours per workday. The suffix

"(pre)" denotes the temporal aspect, indicating data collected to implement yoga practices. The principal output variable, the Venous Clinical Severity Score, was assessed before the yoga intervention (VCSS-PRE) and one month after its initiation (VCSS-1). To ensure methodological rigor, the amassed datasets underwent a randomized allocation into training and testing subsets, maintaining proportions of 70% and 30%, respectively.

The interplay between input and output variables is graphically depicted through a correlation matrix, as illustrated in Fig. 1. Examining the Pearson correlation coefficients reveals discernible patterns. Notably, certain cardiometabolic and vascular health indicators, such as Diabetes Blood pressure (DBP) and Systolic Blood Pressure (SBP), exhibit a strong positive correlation, while the individual's height demonstrates a negative influence on Body Mass Index (BMI). Further scrutiny of the figure highlights that variable RI CA-CIR-PRE-pre and CVIQ_total_pre exert the most pronounced impact on both Venous Clinical Severity Score (VCSS) values. Additionally, it is evident from the analysis that Physical Chronic Fatigue Syndrome (CFS) exerts a more substantial effect than its mental counterpart, particularly concerning VCSS-1. These findings underscore the intricate relationships and varying degrees of influence among the considered variables, providing valuable insights into the dynamics of the observed phenomena.
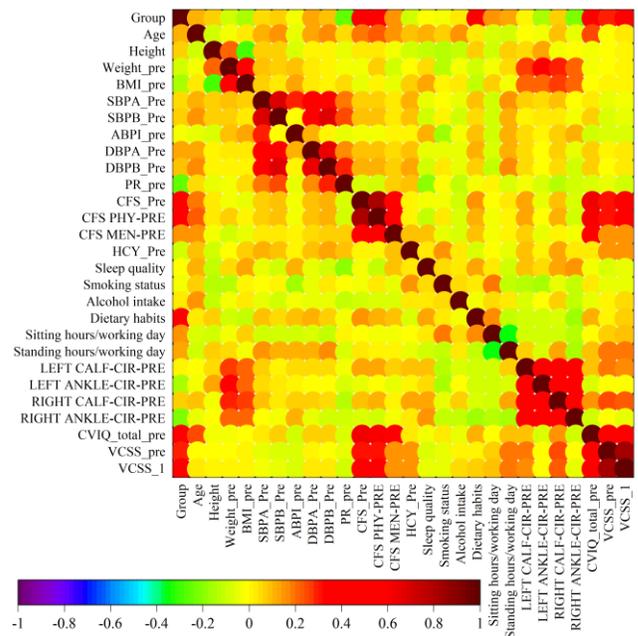

Fig. 1. Correlation matrix to analyze the relationships between input and output variables.

III. RESULTS

A. Evaluation of Models' Applicability

Accuracy is a commonly used statistic in classification issues to assess the overall performance of a model. False Positives (FP) for inaccurate positive forecasts, False Negatives (FN) for wrong negative predictions, and True Positives (TP) for right positive predictions are the four essential components

that it depends on. Accuracy's tendency to favor the majority class, however, may restrict its use in cases with unbalanced data. To alleviate this constraint, three additional assessment measures are often used: F1-Score, Precision, and Recall. In cases when class distributions are unbalanced, these measures provide a more sophisticated evaluation of a model's performance. These metrics are defined through Eq. (28) to Eq. (31). Moreover, it collectively provides a more comprehensive evaluation of a classification model's effectiveness.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{28}$$

$$Precision = \frac{TP}{TP + FP} \tag{29}$$

$$Recall = TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{30}$$

$$F1\_score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{31}$$

B. Convergence Results

This study employed DMO and SAO optimization algorithms to enhance the Support Vector Classification (SVC) model, creating SVDM and SVSA hybrid models. The evaluation of these models utilized a convergence curve based on Accuracy measurements over 150 iterations, revealing a significant improvement in predictive Accuracy (see Fig. 2). In predicting Venous Clinical Severity Score before yoga intervention (VCSS-PRE), both SVDM and SVSA exhibited a substantial increase in Accuracy, reaching peak levels of 0.88 and 0.87 around the 90th iteration. For VCSS-1 prediction, the Accuracy improvement rate was higher, with SVDM and SVSA achieving levels of 0.92 and 0.91, respectively. Notably, SVDM consistently outperformed SVSA in both predictive scenarios, emphasizing its superior ultimate Accuracy. These findings underscore the effectiveness of the optimization algorithms in refining model performance and highlight the comparative advantages of the SVDM hybrid model.
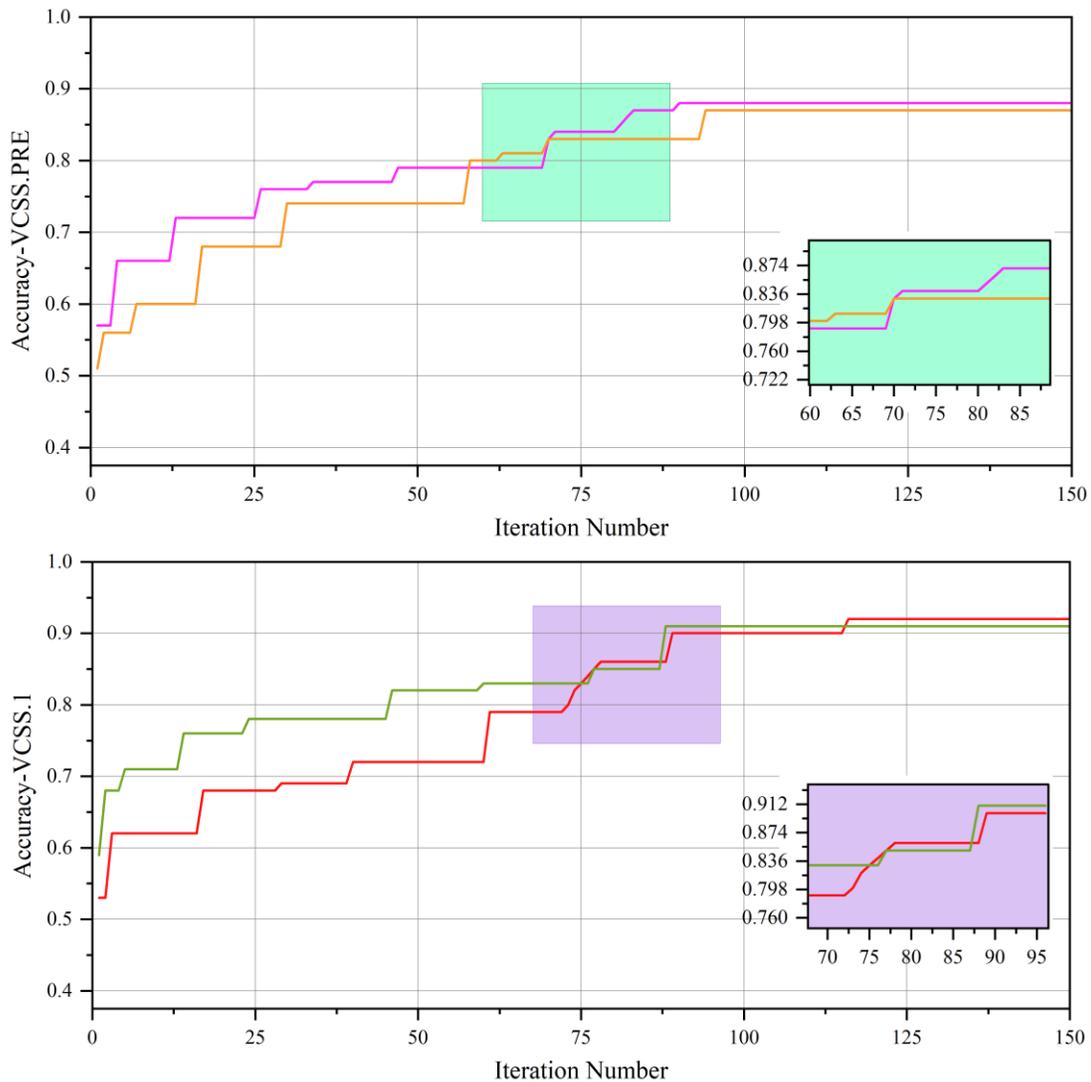

Fig. 2. Convergence curve of hybrid models.

C. Comparing Results of Predictive Models

The primary aim of this study was to introduce three predictive models utilizing a classification approach for anticipating Venous Clinical Severity Score before (VCSS-PRE) and one month after (VCSS-1) yoga practices. Among these models, one employed a Support Vector Classifier (SVC), while the others were developed by optimizing the SVC using Dwarf Mongoose Optimization (DMO) and Smell Agent Optimization (SAO). The performance metrics, including Accuracy, Precision, Recall, and F1-score, for the training and testing phases of these machine learning algorithms are presented in Table I. Notably, for both VCSS-

PRE and VCSS-1 prediction, the metrics during the training phase exceeded those in the testing phase, as visually evident in Fig. 3 (shown as 3D bar plots for all metrics and phases), indicating the models' ineffective training capability. In the case of VCSS-PRE prediction values, the SVDM model demonstrated superior performance, achieving 0.88 for Accuracy and Recall, 0.898 for Precision, and 0.885 for F1_Score. In VCSS-1 estimation, the SVDM model consistently outperformed, recording the highest values across all metrics ($Accuracy = 0.943, Precision = 0.949, Recall = 0.943, and F1 - score = 0.944$).

TABLE I.   RESULT OF PRESENTED MODELS

|	Model	Part	Metric value			
			Accuracy	*Precision*	*Recall*	*F1 _Score*
VCSS-PRE	SVC	*Train*	0.943	0.948	0.943	0.943
		Test	0.667	0.694	0.667	0.677
		All	0.860	0.871	0.860	0.863
	SVDM	*Train*	0.929	0.931	0.9286	0.9287
		Test	0.700	0.7667	0.700	0.7163
		All	0.880	0.898	0.880	0.885
	SVSA	*Train*	0.900	0.918	0.900	0.903
		Test	0.800	0.834	0.800	0.808
		All	0.870	0.892	0.870	0.875
VCSS-1	SVC	*Train*	0.943	0.946	0.943	0.943
		Test	0.800	0.795	0.800	0.793
		All	0.900	0.901	0.900	0.900
	SVDM	*Train*	0.943	0.949	0.943	0.944
		Test	0.867	0.869	0.867	0.866
		All	0.920	0.925	0.920	0.921
	SVSA	*Train*	0.943	0.953	0.943	0.945
		Test	0.833	0.850	0.833	0.835
		All	0.910	0.923	0.910	0.913
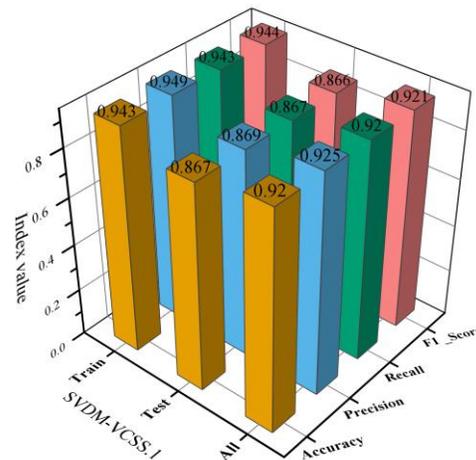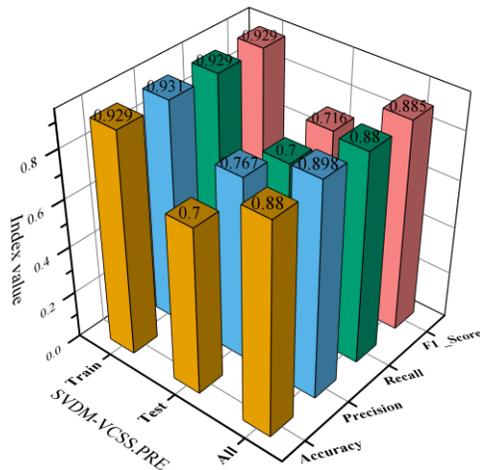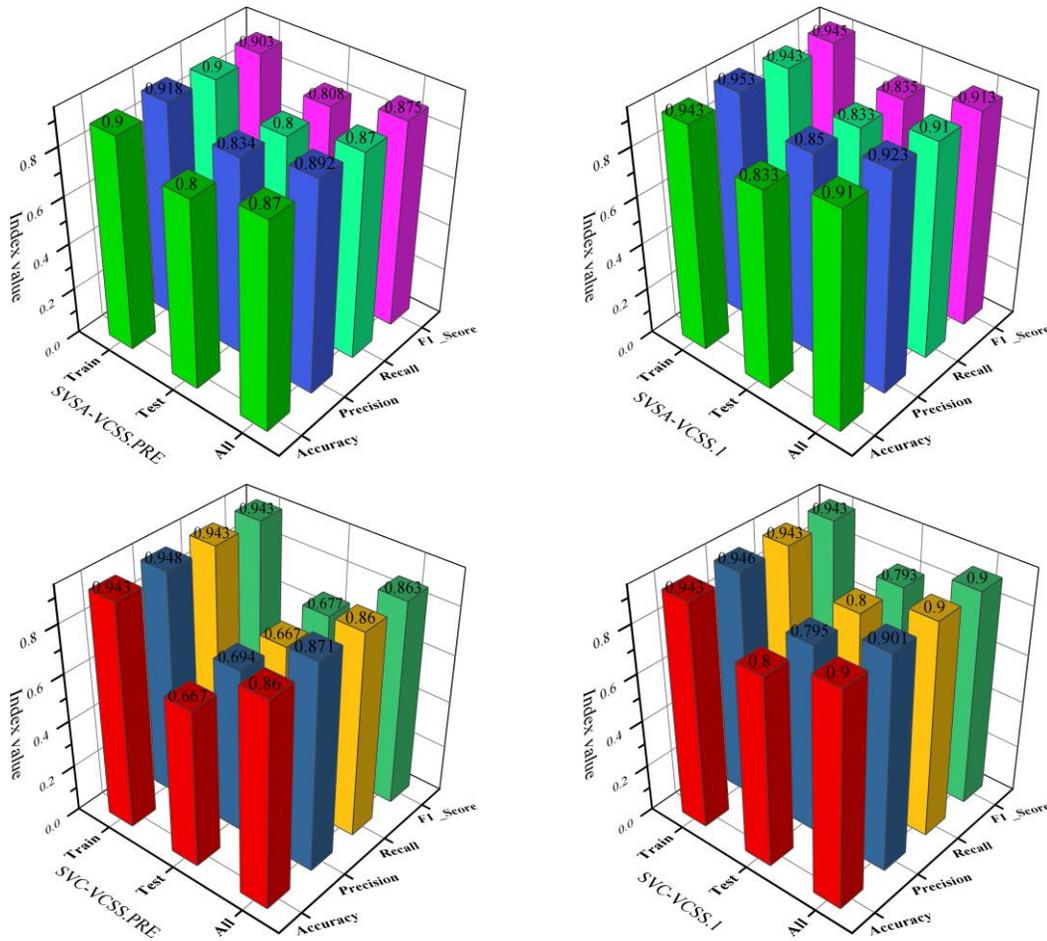
Fig. 3.   3D bar plot to visually assess the performance of the developed models.

The Venous Clinical Severity Score (VCSS) test findings of the 100 samples were used to divide them into four groups after the completion of data processing and a thorough assessment of the models' classification performance in both the training and testing stages. These were divided into four categories: Moderate (11–20), Severe (21–30), Mild (6–10), and Absent (0–5). Tables II and III were created in order to provide a thorough evaluation of the models' categorization effectiveness within each group. These tables provide the Precision, Recall, and F1-score index values—values that are critical for assessing the precision, completeness, and overall accuracy of the models that were generated during the course of the VCSS categories. This granular analysis facilitates a nuanced understanding of the models' performance in distinguishing varying degrees of severity within the studied population, contributing valuable insights to the overall assessment of their predictive capabilities.

1) Precision

a) VCSS-PRE: The SVSA model demonstrated the greatest accuracy values in the Mild and Severe categories, with scores of 0.881 and 1.000, respectively. On the other hand, in the Absent group, the SVDM model reached its maximum accuracy value of 0.643. Notably, the SVC model

outperformed the other models for the Moderate category, earning an accuracy score of 1.000.

b) VCSS-1: The SVSA model showcased superior Precision across the Mild, Moderate, and Severe categories, securing impressive scores of 0.939, 0.978, and 1.000, respectively. In contrast, the Absent group saw the SVC model achieving its maximum precision value of 0.867.

2) Recall

a) VCSS-PRE: The SVDM model excelled with the highest scores in the Mild (0.905), Moderate (0.864), and Severe (1.000) groups. Contrastingly, the SVSA model delivered an outstanding performance for the Absent group, attaining the top recall score of 0.917.

b) VCSS-1: Attaining Recall values of 1.000 and 0.892, respectively, the SVDM model demonstrated exceptional performance in the Absent and mild categories. For the Moderate and Severe groups, the SVSA model also produced maximum recall values of 0.957 and 1.000.

3) F1-score

a) VCSS-PRE: A high F1 score indicates that the model is able to discriminate between accurately detecting positive instances (Precision) and include all true positive cases (Recall). The SVDM model performed better than all other

models in every category, with the greatest F1-scores in the Mild (0.884), Moderate (0.927), and Severe (1.000) groups. Additionally, for the Absent group, the SVSA model reached its maximum F1-Score value of 0.733.

b) VCSS-1: The SVDM model performed better in the Absent and Mild categories, with F1-Score values of 0.903 and 0.892, respectively. Furthermore, the SVSA model achieved remarkable ratings of 0.968 and 1.000 in the Moderate and Severe categories, outperforming other models.

The actual count of samples categorized as Absent, Mild, Moderate, and Severe was 12, 42, 44, and 2, respectively, for VCSS-PRE and 14, 37, 47, and 2 for VCSS-1 values. Fig. 4 visually presents these categories, offering a 3D walls-based

comparison for measurements and classification model outcomes. In the context of VCSS-PRE, the SVDM model demonstrated superior accuracy, correctly classifying individuals into the Mild, Moderate, and Severe groups, identifying 38, 38, and 2 individuals accurately, respectively. The SVSA model outperformed other models in the Absent category, accurately classifying 11 individuals. Turning to VCSS-1 values, the SVDM model maintained its Accuracy, correctly classifying individuals in the Absent, Mild, and Severe groups, identifying 14, 33, and 2 individuals accurately, respectively. Notably, in the Moderate category, the SVSA model outperformed other models by accurately classifying 45 individuals.

TABLE II.   EVALUATION INDEXES OF THE DEVELOPED MODELS' PERFORMANCE BASED ON GRADES VCSS-PRE

Model	Grade	Metric value		
		Precision	*Recall*	*F1 − score*
SVC	Absent	0.643	0.750	0.692
	Mild	0.822	0.881	0.851
	Moderate	0.974	0.864	0.916
	Severe	1.000	1.000	1.000
SVDM	Absent	0.625	0.833	0.714
	Mild	0.864	0.905	0.884
	Moderate	1.000	0.864	0.927
	Severe	1.000	1.000	1.000
SVSA	Absent	0.611	0.917	0.733
	Mild	0.881	0.881	0.881
	Moderate	0.974	0.841	0.902
	Severe	1.000	1.000	1.000

TABLE III.   EVALUATION INDEXES OF THE DEVELOPED MODELS' PERFORMANCE BASED ON GRADES VCSS-1

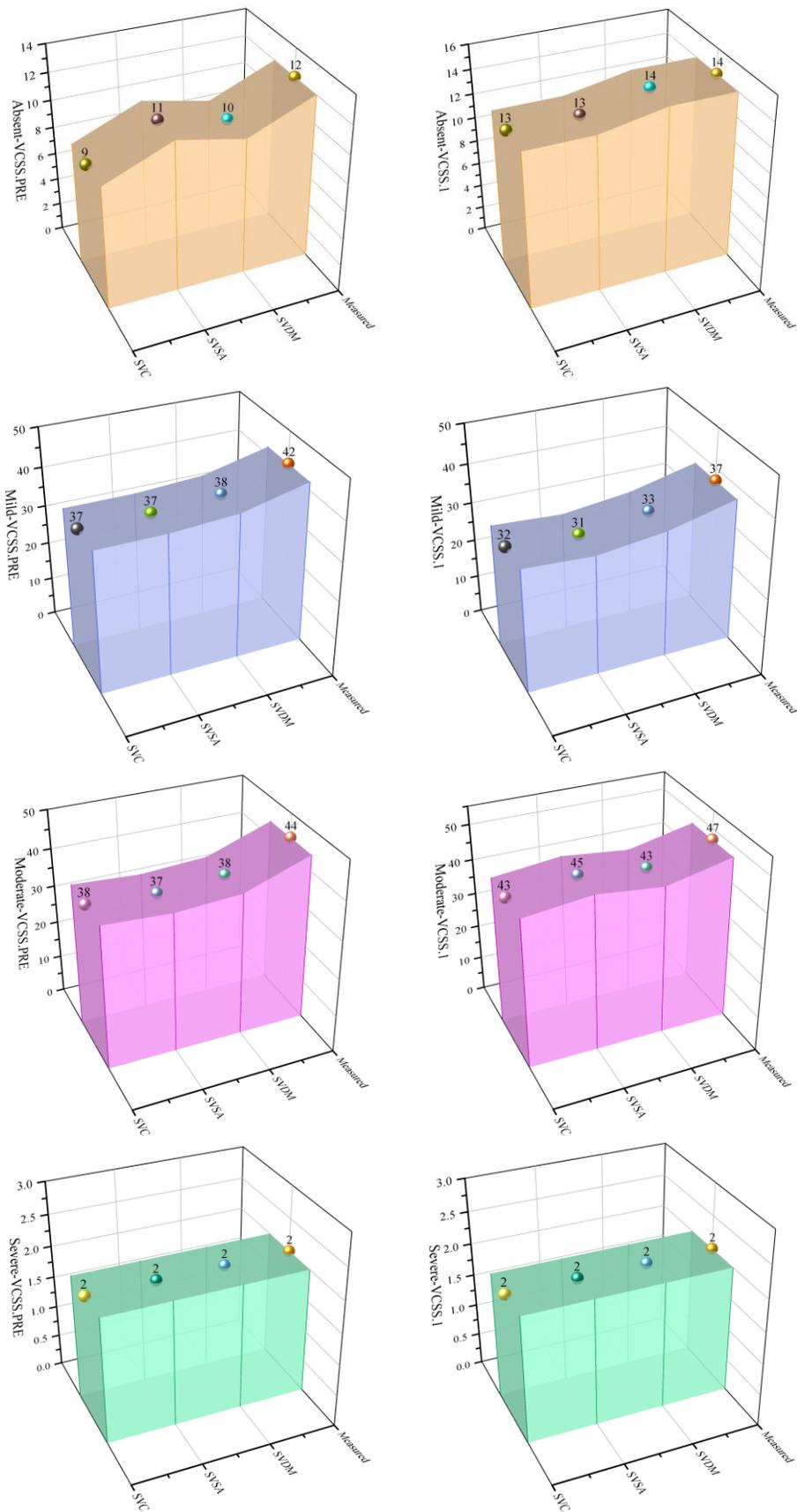| Model	Grade	Metric value		
		Precision	*Recall*	*F1 − score*
SVC	Absent	0.867	0.929	0.897
	Mild	0.865	0.865	0.865
	Moderate	0.935	0.915	0.925
	Severe	1.000	1.000	1.000
SVDM	Absent	0.824	1.000	0.903
	Mild	0.892	0.892	0.892
	Moderate	0.977	0.915	0.945
	Severe	1.000	1.000	1.000
SVSA	Absent	0.684	0.929	0.788
	Mild	0.939	0.838	0.886
	Moderate	0.978	0.957	0.968
	Severe	1.000	1.000	1.000

Fig. 4.   3D walls for the difference between measured and predicted values.

Understanding the confusion matrix in Fig. 5 may help with correctly classifying people into the appropriate groups and identifying those who are misclassified into other groups. In reference to VCSS-PRE data, the SVDM model accurately classified 2, 38, 38, and 10 individuals into the Severe, Moderate, Mild, and Absent classifications, respectively; only 14 pupils were misclassified. But, the SVC and SVSA models incorrectly categorized 16 and 15, respectively, of the individuals. The two optimized models showed that misclassifications mostly occurred across adjacent categories.

For example, four individuals from SVSA and SVC were incorrectly classified as belonging to the Mild group instead of the Absent category. Twelve people were misclassified by the SVC model, which accurately classified 2, 43, 32, and 13 people into the Severe, Moderate, Mild, and Absent categories, respectively, based on VCSS-1 scores. The SVSA and SVDM models, on the other hand, incorrectly categorized 11 and 10 people, respectively. According to the SVSA model, five kids were mistakenly assigned to the Mild category rather than the Absent group.
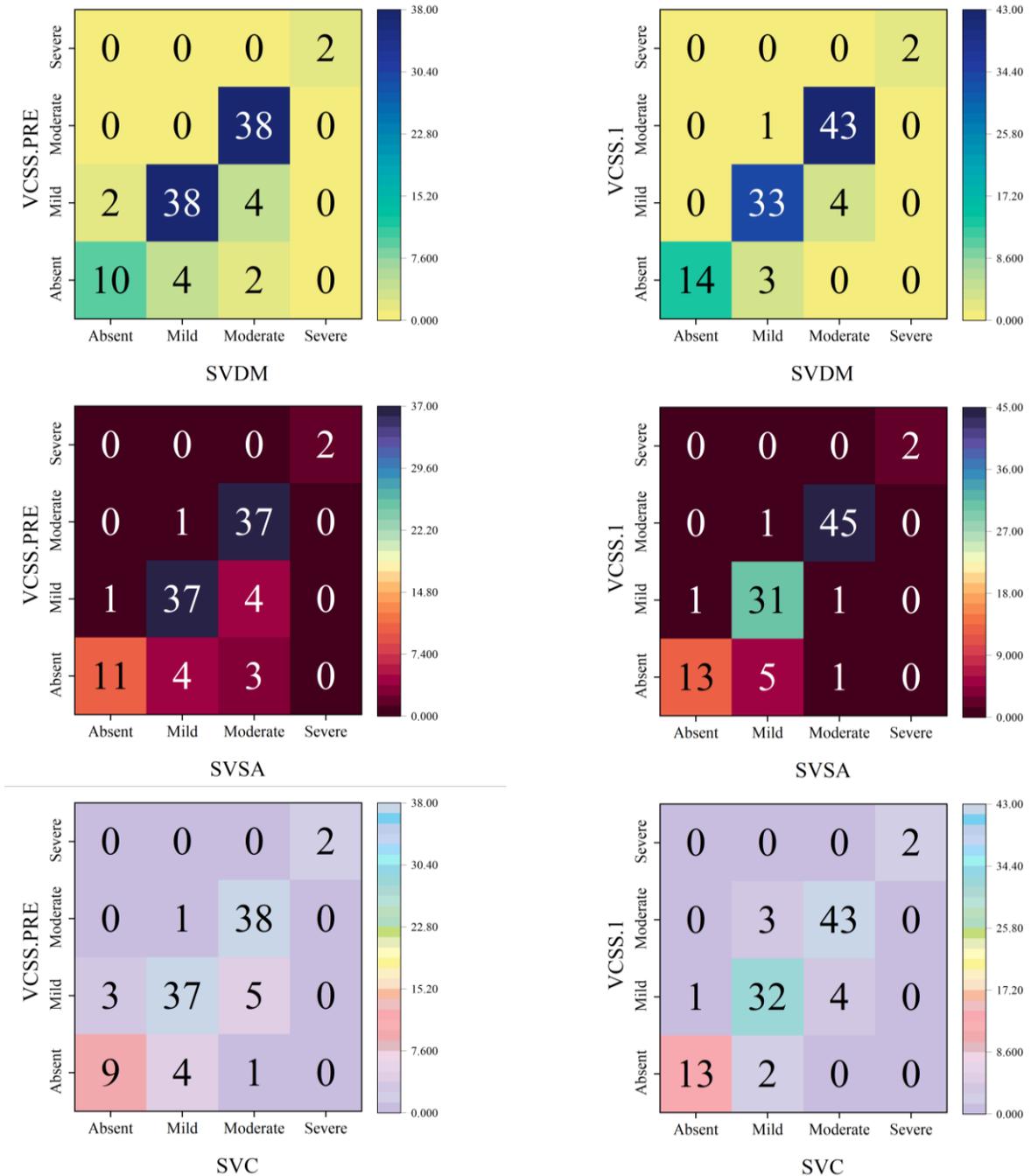

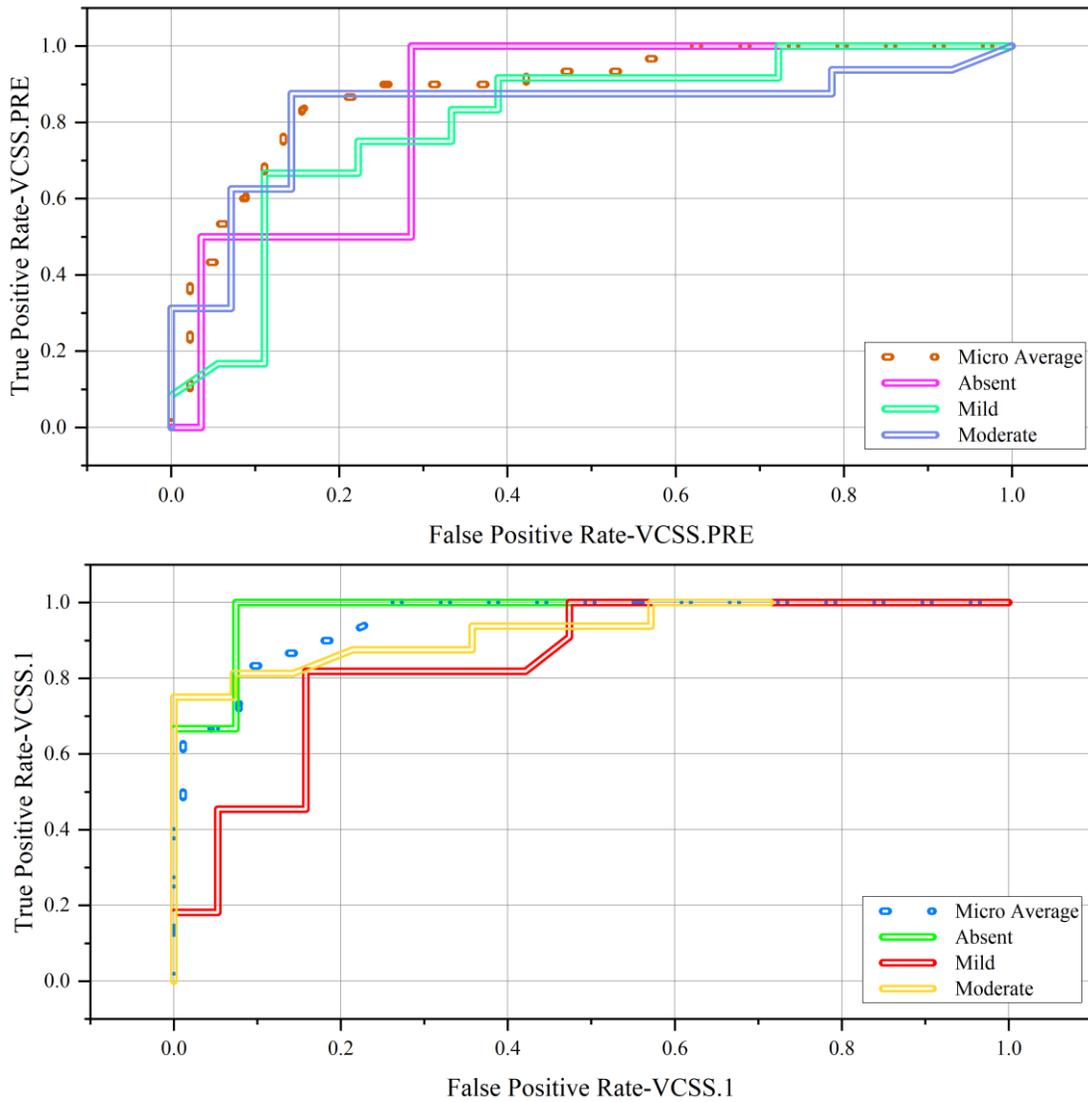Fig. 5.   Confusion matrix for the accuracy of each model.

Fig. 6. The ROC curve for comparison of the SVDM model between various categories.

By employing the Receiver Operating Characteristic (ROC) curve, the evaluation seeks to discern the equilibrium between the True Positive (TP) and False Positive (FP) rates, complemented by the computation of the Area Under the ROC Curve (AUC). A higher AUC signifies a more controlled increase in the FP rate compared to a substantial rise in the TP rate for each adjustment of the predicted probability threshold. An ideal discrimination test is characterized by a ROC plot reaching the upper-left corner, signifying 100% sensitivity and specificity. Fig. 6, which depicts ROC curves for the optimal SVDM model in classifying samples across two VCSS periods, illustrates that in VCSS-PRE, the AUC related to the Moderate group exceeded other categories and exhibited a more pronounced inclination towards the left-top side of the diagram. In the case of VCSS-1, the AUC for the Moderate and Absent groups surpassed that of the Mild curve.

D. Sensitivity Analyses

1) SHAP: SHAP (SHapley Additive exPlanations) is an algorithm used for interpreting machine learning models. It assigns Shapley values to each feature, indicating their individual contributions to model predictions. Derived from cooperative game theory, Shapley values ensure a fair distribution of the model's output among features by considering all possible feature combinations [42], [43]. This approach provides both local and global interpretability, explaining predictions for specific instances and revealing overall model behavior. SHAP values can be visualized through various plots, aiding in the understanding of complex models and building trust by uncovering the factors influencing predictions.

Fig. 7 shows the effect of inputs on the output of the model. Based on the analysis, it was observed that CFS_Pre had the highest impact on the model output and Group had the lowest impact in all four classifications.
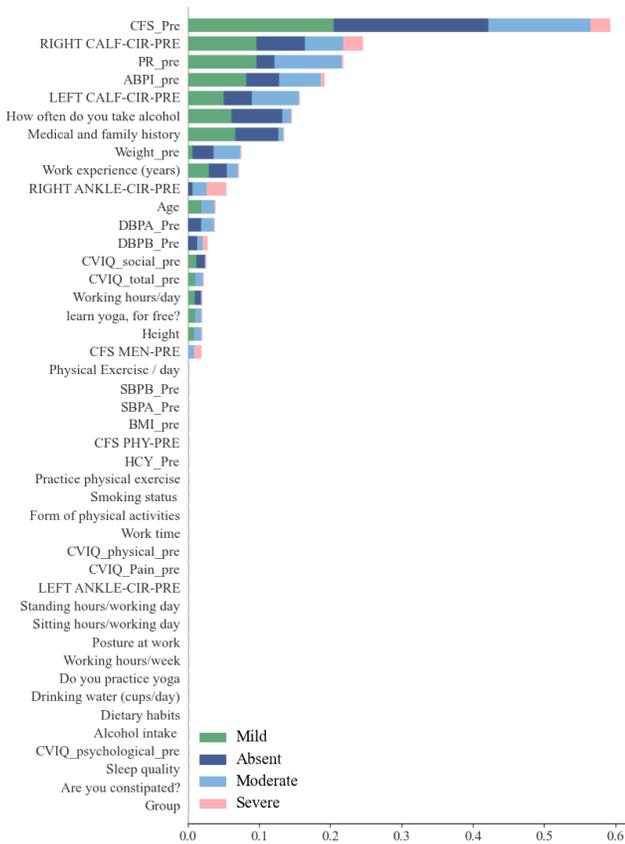
Fig. 7.   Impact of input variables on model's output.

IV.   DISCUSSION

The study has several limitations that should be considered in interpreting its findings. Firstly, the reliance on a sample size of 100 participants may restrict the generalizability of the results to broader populations. Future research endeavors should prioritize larger and more diverse samples to enhance external validity and ensure a representative study cohort. Additionally, the study's exploration of the duration of non-pharmacological interventions, particularly yoga, was somewhat limited. A more in-depth investigation into longer intervention periods could provide valuable insights into the sustainability of effects and potential long-term benefits. Furthermore, the study predominantly focused on yoga as a non-pharmacological intervention, potentially limiting the breadth of its applicability. Future research could benefit from investigating the comparative effectiveness of various non-pharmacological interventions, taking into consideration individual preferences and adherence rates. The study's reliance on quantitative outcome measures, while valuable, might not fully capture the nuanced impact of interventions on participants' daily lives and overall well-being. Incorporating qualitative assessments and patient-reported outcomes in future studies could provide a more comprehensive understanding of the holistic effects of these interventions.

On the other hand, the study's findings offer promising applications in clinical settings and beyond. The optimization of non-pharmacological interventions using machine learning algorithms, as demonstrated in the study, suggests potential effectiveness in managing CVI. This could encourage healthcare practitioners to consider integrating such interventions into comprehensive patient care plans, especially for individuals with varying levels of CVI severity.

Moreover, the study contributes to the evolving landscape of personalized medicine by showcasing the potential of machine learning models in tailoring interventions based on individual CVI profiles. This has implications for future applications, with the prospect of refining algorithms for more precise and personalized treatment recommendations. The findings may also have relevance in healthcare policy discussions, emphasizing the value of non-pharmacological approaches in addressing CVI. Policymakers could consider strategies to promote the integration of these interventions within healthcare systems, potentially leading to cost-effective and patient-centered care.

V.   CONCLUSION

This investigation navigates the crossroads of technology, healthcare, and preventive strategies, delving into the potential of non-pharmacological interventions, notably yoga, to alleviate the urgency associated with Chronic Venous Insufficiency (CVI). Particularly, the study addresses the impact of such interventions during periods of heightened stress and sedentary lifestyles. The research unfolds avenues for predictive modeling and precision medicine by demonstrating the fusion of machine learning algorithms with healthcare. Leveraging a data-driven approach across a sample size of 100, the introduction of Support Vector Classification (SVC) models optimized with Dwarf Mongoose Optimization (DMO) and Smell Agent Optimization (SAO) provides valuable insights into the classification of CVI severity levels. Applying DMO and SAO optimization techniques to the SVC model resulted in a significant improvement in accuracy for VCSS-PRE values, with increases of 2% and 1%, respectively. As the 100 individuals were classified according to their circumstances, the DMO's remarkable capacity to improve classification accuracy was made clear. In particular, the SVDM model correctly categorized most people with an astounding accuracy rate of 94.3%, whereas the SVSA and SVC models incorrectly classified 15% and 16% of all people, respectively. When it comes to VCSS-1 values, the introduction of DMO and SAO optimization techniques to the SVC model improved Precision by 2.4% and 2.2%, respectively. With an accuracy rate of just 80%, the SVC model correctly classified the fewest individuals, whereas the SVSA and SVDM models had classification rates of 89% and 90%, respectively. Further investigations into non-pharmacological interventions and CVI could contribute to the body of knowledge by implementing a longitudinal design to monitor the long-term impact, ensuring a diverse range of participants to enhance the generalizability of findings, and conducting comparative analyses of interventions such as mindfulness and yoga. By incorporating patient-reported outcomes and investigating the various factors that impact adherence, a comprehensive understanding can be achieved. The integration of sophisticated imaging methodologies will provide impartial assessments of the advancement of CVI, whereas health economics evaluations can scrutinize cost-effectiveness. Collaboration with healthcare professionals and

mechanistic investigation can enhance understanding of intervention pathways and promote the adoption of multidisciplinary approaches. Ethical considerations are of the utmost importance, encompassing participant safety and informed consent.

COMPETING OF INTERESTS

The authors declare no competing of interests.

AUTHORSHIP CONTRIBUTION STATEMENT

Xiao Du: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

DATA AVAILABILITY

The author does not have permission to share data.

DECLARATIONS

Not applicable.

REFERENCES

[1] C. Liddy, J. Singh, W. Hogg, S. Dahrouge, and M. Taljaard, "Comparison of primary care models in the prevention of cardiovascular disease-a cross sectional study," BMC Fam Pract, vol. 12, pp. 1–10, 2011.

[2] B. Bottazzi, E. Riboli, and A. Mantovani, "Aging, inflammation and cancer," in Seminars in immunology, Elsevier, 2018, pp. 74–82.

[3] N. G. Vallianou, T. Stratigou, and S. Tsagarakis, "Microbiome and diabetes: where are we now?," Diabetes Res Clin Pract, vol. 146, pp. 111–118, 2018.

[4] S. Onida and A. H. Davies, "Predicted burden of venous disease," Phlebology, vol. 31, no. 1_suppl, pp. 74–79, 2016.

[5] [5] L. Robertson, C. and Evans, and F. G. R. Fowkes, "Epidemiology of chronic venous disease," Phlebology, vol. 23, no. 3, pp. 103–111, 2008.

[6] J. L. Beebe-Dimmer, J. R. Pfeifer, J. S. Engle, and D. Schottenfeld, "The epidemiology of chronic venous insufficiency and varicose veins," Ann Epidemiol, vol. 15, no. 3, pp. 175–184, 2005.

[7] N. C. G. C. UK, "Varicose Veins in the Legs: The Diagnosis and Management of Varicose Veins," 2013.

[8] M.-L. Kuet, T. R. A. Lane, M. A. Anwar, and A. H. Davies, "Comparison of disease-specific quality of life tools in patients with chronic venous disease," Phlebology, vol. 29, no. 10, pp. 648–653, 2014.

[9] J. El-Sheikha, "A multilevel regression of patient-reported outcome measures after varicose vein treatment in England," Phlebology, vol. 31, no. 6, pp. 421–429, 2016.

[10] E. Rabe and F. Pannier, "Societal costs of chronic venous disease in CEAP C4, C5, C6 disease," Phlebology, vol. 25, no. 1_suppl, pp. 64–67, 2010.

[11] A. Couce and J. Blazquez, "Side effects of antibiotics on genetic variability," FEMS Microbiol Rev, vol. 33, no. 3, pp. 531–538, 2009.

[12] B. A. Cunha, "Antibiotic side effects," Medical Clinics of North America, vol. 85, no. 1, pp. 149–185, 2001.

[13] H. Cramer, L. Ward, A. Steel, R. Lauche, G. Dobos, and Y. Zhang, "Prevalence, patterns, and predictors of yoga use: results of a US nationally representative survey," Am J Prev Med, vol. 50, no. 2, pp. 230–235, 2016.

[14] D. Neumark-Sztainer, A. W. Watts, and S. Rydell, "Yoga and body image: How do young adults practicing yoga describe its impact on their body image?," Body Image, vol. 27, pp. 156–168, 2018.

[15] E. Halliwell, K. Dawson, and S. Burkey, "A randomized experimental evaluation of a yoga-based body image intervention," Body Image, vol. 28, pp. 119–127, 2019.

[16] R. R. Guddeti, G. Dang, M. A. Williams, and V. M. Alla, "Role of yoga in cardiac disease and rehabilitation," J Cardiopulm Rehabil Prev, vol. 39, no. 3, pp. 146–152, 2019.

[17] G. Sathyanarayanan, A. Vengadavaradan, and B. Bharadwaj, "Role of yoga and mindfulness in severe mental illnesses: A narrative review," Int J Yoga, vol. 12, no. 1, p. 3, 2019.

[18] S. Patil, A. Pawar, A. Peshave, A. N. Ansari, and A. Navada, "Yoga tutor visualization and analysis using SURF algorithm," in 2011 IEEE control and system graduate research colloquium, IEEE, 2011, pp. 43–46.

[19] H.-T. Chen, Y.-Z. He, C.-C. Hsu, C.-L. Chou, S.-Y. Lee, and B.-S. P. Lin, "Yoga posture recognition for self-training," in MultiMedia Modeling: 20th Anniversary International Conference, MMM 2014, Dublin, Ireland, January 6-10, 2014, Proceedings, Part I 20, Springer, 2014, pp. 496–505.

[20] H.-T. Chen, Y.-Z. He, C.-L. Chou, S.-Y. Lee, B.-S. P. Lin, and J.-Y. Yu, "Computer-assisted self-training system for sports exercise using kinects," in 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), IEEE, 2013, pp. 1–4.

[21] M. B. Schure, J. Christopher, and S. Christopher, "Mind–body medicine and the art of self-care: teaching mindfulness to counseling students through yoga, meditation, and qigong," Journal of Counseling & Development, vol. 86, no. 1, pp. 47–56, 2008.

[22] R. Shad, J. P. Cunningham, E. A. Ashley, C. P. Langlotz, and W. Hiesinger, "Designing clinically translatable artificial intelligence systems for high-dimensional medical imaging," Nat Mach Intell, vol. 3, no. 11, pp. 929–935, 2021.

[23] N. Komal Kumar, D. Vigneswari, M. Vamsi Krishna, and G. V Phanindra Reddy, "An optimized random forest classifier for diabetes mellitus," in Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018, Volume 2, Springer, 2019, pp. 765–773.

[24] D. Vigneswari, N. K. Kumar, V. G. Raj, A. Gugan, and S. R. Vikash, "Machine learning tree classifiers in predicting diabetes mellitus," in 2019 5th international conference on advanced computing & communication systems (ICACCS), IEEE, 2019, pp. 84–87.

[25] V. Mareeswari, R. Saranya, R. Mahalakshmi, and E. Preethi, "Prediction of diabetes using data mining techniques," Res J Pharm Technol, vol. 10, no. 4, pp. 1098–1104, 2017.

[26] T. Sudhakar, J. B. Janney, D. Haritha, M. J. Sahaya, and V. Parvathy, "Automatic Detection and Classification of Brain Tumor using Image Processing Techniques," Res J Pharm Technol, vol. 10, no. 11, pp. 3692–3696, 2017.

[27] T. Nafee et al., "Machine learning to predict venous thrombosis in acutely ill medical patients," Res Pract Thromb Haemost, vol. 4, no. 2, pp. 230–237, 2020.

[28] L. Ryan et al., "A machine learning approach to predict deep venous thrombosis among hospitalized patients," Clinical and Applied Thrombosis/Hemostasis, vol. 27, p. 1076029621991185, 2021.

[29] N. K. Kumar, G. S. Sindhu, D. K. Prashanthi, and A. S. Sulthana, "Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers," in 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 15–21. doi: 10.1109/ICACCS48705.2020.9074183.

[30] V. Vapnik, "Statistical Learning Theory. New York: John Willey & Sons," Inc, 1998.

[31] S. Maldonado, J. Pérez, R. Weber, and M. Labbé, "Feature selection for support vector machines via mixed integer linear programming," Inf Sci (N Y), vol. 279, pp. 163–175, 2014.

[32] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," ACM transactions on intelligent systems and technology (TIST), vol. 2, no. 3, pp. 1–27, 2011.

[33] M. Aydogdu and M. Firat, "Estimation of failure rate in water distribution network using fuzzy clustering and LS-SVM methods," Water resources management, vol. 29, pp. 1575–1590, 2015.

[34] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.," 2022.

[35] L. B. Buck, "Unraveling the sense of smell," Les Prix Nobel. The Nobel Prizes, vol. 2004, pp. 267–283, 2004.

[36] E. Sakalli, D. Temirbekov, E. Bayri, E. E. Alis, S. C. Erdurak, and M. Bayraktaroglu, "Ear nose throat-related symptoms with a focus on loss of smell and/or taste in COVID-19 patients," Am J Otolaryngol, vol. 41, no. 6, p. 102622, 2020.

[37] R. Axel, "Scents and sensibility: a molecular logic of olfactory perception (Nobel lecture)," Angewandte Chemie International Edition, vol. 44, no. 38, pp. 6110–6127, 2005.

[38] S. Chapman and T. G. Cowling, The mathematical theory of non-uniform gases: an account of the kinetic theory of viscosity, thermal conduction and diffusion in gases. Cambridge university press, 1990.

[39] M. Abdechiri, M. R. Meybodi, and H. Bahrami, "Gases Brownian motion optimization: an algorithm for optimization (GBMO)," Appl Soft Comput, vol. 13, no. 5, pp. 2932–2946, 2013.

[40] J. O. Agushaka, A. E. Ezugwu, and L. Abualigah, "Dwarf mongoose optimization algorithm," Comput Methods Appl Mech Eng, vol. 391, p. 114570, 2022.

[41] U. Yamuna, V. Majumdar, and A. A. Saoji, "Effect of Yoga on homocysteine level, symptomatology and quality of life in industrial workers with Chronic Venous Insufficiency: Study protocol for a randomized controlled trial," Adv Integr Med, vol. 9, no. 2, pp. 119–125, 2022.

[42] I. U. Ekanayake, D. P. P. Meddage, and U. Rathnayake, "A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP)," Case Studies in Construction Materials, vol. 16, p. e01059, 2022.

[43] Y. Wu and Y. Zhou, "Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete," Constr Build Mater, vol. 330, p. 127298, 2022.