



# Optimized feature selection for enhanced accuracy in knee osteoarthritis detection and severity classification with machine learning



Anandh Sam Chandra Bose<sup>a</sup>, Srinivasan C<sup>b,\*</sup>, Immaculate Joy S<sup>b</sup>

<sup>a</sup> Department of Industrial Engineering, College of Applied Sciences, AlMaarefa University, Kingdom of Saudi Arabia

<sup>b</sup> Department of ECE, Saveetha Engineering College, Chennai, Tamil Nadu, India

## ARTICLE INFO

### Keywords:

Knee osteoarthritis  
Kaggle  
Optimization  
Feature selection  
Prediction  
Images

## ABSTRACT

Knee Osteoarthritis (KOA) is a widespread and debilitating orthopaedic condition that necessitates accurate diagnostic tools. Machine Learning (ML) models have been utilized for KOA detection, whereas there is an ongoing need to enhance the reliability of these models. We believe that precise Feature Selection (FS) plays a crucial role in improving the performance of ML models. This research delves into the pivotal role of FS in enhancing the accuracy and dependability of ML models used in KOA detection and severity classification. The data were obtained from Kaggle, representing various grades of KOA. We employ a Convolutional Neural Network (CNN) model to extract features from medical imaging data. Utilizing advanced techniques such as Particle Swarm Optimization (PSO) and Genetic Bee Colony (GBC), we systematically identify relevant features to enhance our ML models. Comparative analyses between models trained with optimized features and those using direct CNN features reveal significant enhancements in accuracy, specificity, sensitivity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) across multiple ML algorithms including Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), and Random Forest (RF). In binary classification tasks, GBC-selected features achieve highest metrics with 99.15% accuracy, 99.35% sensitivity, 98.89% specificity, 99.14% PPV, 99.16% NPV, and 99.25% F-Measure. In multiclass classification, GBC features combined with RF demonstrate high accuracy (98.91%), sensitivity (98.91%), specificity (98.90%), PPV (99.13%), NPV (98.63%), and F-Measure (99.02%). These findings underscore the efficacy of feature optimization in enhancing ML model in precise KOA diagnosis and advancing patient care in orthopaedic medicine.

## 1. Introduction

KOA is a progressive degenerative condition affecting the knee joint, impacting its three compartments (lateral, medial, and patella-femoral) [1]. Typically, it develops slowly throughout 10 to 15 years, often stemming from wear and tear on the joint and occasionally from infections causing damage to the joint cavity. This results in discomforts like limited mobility, joint pain, and swelling [2,3]. Osteoarthritis (OA) can impact various joints in the human body, although it tends to affect weight-bearing joints, such as the knees and hips, more significantly. Furthermore, KOA primarily targets individuals aged 55 and above, with a higher occurrence rate observed among those over the age of 65 [4]. Experts project that by the year 2050, approximately 130 million individuals across the globe will grapple with KOA [5]. Nevertheless, timely identification and intervention can play a crucial role in mitigating the disease's advancement and improving the overall well-being

of those afflicted by it.

Diagnosing and treating KOA is challenging due to its complexity and numerous risk factors, including hormonal status, age, gender, and Body Mass Index (BMI) [6]. Furthermore, numerous medical, environmental, and biological factors, some of which can be modified and others which cannot, contribute to its onset and advancement. In severe instances, individuals with these risk factors may necessitate undergoing total knee replacement surgery. At present, the existing treatments for KOA predominantly revolve around behavioural interventions such as weight management, physical activity, and the strengthening of joint muscles [7]. These interventions can offer temporary relief from pain while also decelerating the progression of the condition. Radiographs (X-rays) are commonly used to diagnose and assess KOA, being the preferred method due to their cost-effectiveness, safety, widespread availability, and quick results [8]. Radiologists commonly search for particular pathological characteristics in knee X-ray images, specifically focusing on signs such

\* Corresponding author.

E-mail address: [csrcinivasan.1983@gmail.com](mailto:csrcinivasan.1983@gmail.com) (C. Srinivasan).

as Joint Space Narrowing (JSN) and the development of osteophytes. These indicators are also utilized in assessing the degree of KOA severity through the Kellgren–Lawrence (KL) grading system, which categorizes it into five grades, ranging from grade 0 to grade 4 [9]. Grade 0 indicates healthy joints with no signs of KOA. Grade 1 suggests uncertain KOA, with possible osteophytes and questionable JSN. Grade 2 indicates mild OA, characterized by clear osteophytes and possible JSN. Grade 3 signifies moderate OA, with multiple osteophytes, JSN, and sclerosis. Lastly, Grade 4 represents severe OA, marked by substantial osteophytes, JSN, and severe sclerosis. Considering the limited availability of radiologists, particularly in rural areas, and the time-intensive nature of knee X-ray analysis, there is a growing demand for fully automated knee severity classification systems. DL and ML techniques have gained prominence in the field of medical imaging, addressing challenges related to classification, detection, and various other aspects without relying on the expertise of a radiologist [6,10,11]. Consequently, several Computer-Aided Diagnosis (CAD)-based approaches in medical imaging have been proposed to detect and assess KOA. Nevertheless, there remains a need to enhance the reliability of existing models. To enhance the accuracy of KOA detection, the selection of features supplied to the model holds paramount importance. To address this, we have employed optimized FS techniques.

Unlike traditional approaches that directly feed extracted features into ML models, we employ a sequential feature processing pipeline. Initially, features are extracted from preprocessed images using CNNs. Subsequently, advanced optimization techniques like PSO and GBC are applied to select the most relevant subset of features. This sequential processing ensures that only the most discriminative features are utilized by the ML models, thereby enhancing their performance. We conduct a comprehensive evaluation of multiple ML algorithms using the optimized feature sets. Through this systematic comparison, we identify the best combination of optimization and ML algorithms for yielding the highest accuracy rates. The novelty of the research lies in utilizing CNN to extract features and employing bio-optimized algorithms such as PSO and GBC for selecting the most significant features. This approach is further enhanced by integrating PSO and GBC with multiple ML models to determine the optimal combination of FS techniques and ML models for the accurate detection and grading of KOA.

## 2. Literature survey

This section provides a comprehensive overview of the literature review conducted as part of this research. Below are some recent journals on medical image analysis that have been explored, classified, and summarized. For improving accuracy, some research focuses on feature extraction and selection, while others focus on ML and DL model design. We have categorized the literature survey based on the priority given by the authors to improve accuracy.

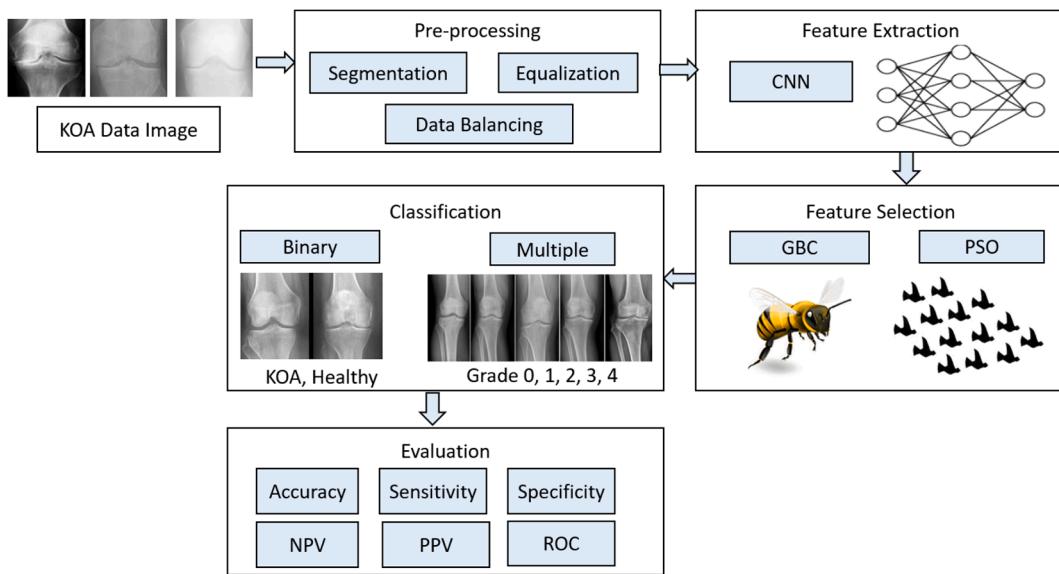
### 2.1. Feature extraction and selection

In the paper [12], a novel approach is proposed for KOA detection and severity classification using a customized CenterNet with a pixel-wise voting scheme for automatic Feature Extraction (FE). The model prioritizes the most appropriate features based on accurate localization findings and applies a weighted pixel-wise voting approach to improve bounding box forecasting accuracy. In addition, the research uses the idea of knowledge distillation, which transfers knowledge from a complicated network to a simpler one without raising computing costs. This method allows for accurate KOA identification and categorization. The suggested framework extends CenterNet by using a basic DenseNet-201 for FE. The model is trained and tested on datasets, demonstrating superior performance on testing data. In the study [13], the goal is to develop a robust FS methodology capable of handling multidimensional datasets and addressing the limitations of existing FS techniques in identifying significant risk factors for KOA diagnosis. The suggested

fuzzy ensemble FS technique integrates outcomes from diverse FS algorithms, encompassing filter, wrapper, and embedded methods, employing fuzzy logic. Thorough experimentation involving multiple FS algorithms and established ML models underscores the efficacy of this approach. It attains an impressive classification accuracy of 73.55 % with the top-performing model being the RF classifier, employing a set of twenty-one chosen risk factors. Additionally, the research includes an analysis to measure the influence of selected features on the model's output, enriching our comprehension of the decision-making process.

In the study [14], the objective is to create an effective model for early-stage OA diagnosis in knee X-ray images. The proposed approach combines advanced DL-based CNN and ML techniques. A novel transfer learning-based feature engineering technique, CRK (CNN+RF+KNN), is introduced for high-performance OA detection. This approach uses a 2D-CNN to extract spatial features from X-ray images, which are then input into RF and KNN techniques to create a probabilistic feature set. The proposed model achieves a remarkable accuracy score for OA prediction, and extensive experiments include hyper-parameter optimization and k-fold cross-validation. Knee-joint Vibration Arthrometry Graph (VAG) signal analysis is important during the initial pathological screening of the knee joint, providing an effective non-invasive technique for KOA detection. The study [15] provides a KOA pathology screening approach that uses Time-Domain Multidimensional Fusion Features (TDMFF) in combination with the RF classifier to increase the diagnostic accuracy of KOA. This method incorporates feature fusion to establish a model for time-domain multidimensional fusion features, which effectively captures the fusion feature of VAG signals. Experimental validation of this methodology is conducted using both normal and abnormal VAG signals from test subjects. The results of KOA screening demonstrate impressive results. The study [16] introduces an integrated approach that leverages ML and FE to enhance the precision of radiographic image-based KOA stage diagnosis and classification. The study used FE approaches such Histogram of Oriented Gradients with LDA and Min-Max scale for categorization. The study tested six ML classifiers that had been fine-tuned with the use of GridSearchCV's hyperparameter tuning methods. In addition to the high-accuracy models, an ensemble model was developed to improve accuracy while simultaneously reducing the likelihood of overfitting. The XGBoost and the ensemble model demonstrated remarkable accuracy in differentiating between healthy and KOA, emerging as the most effective for multiclass classification.

Research [17] describes the development of a KOA diagnostic system utilizing the Canonical correlation estimation technique along with Gabor and Tamura factors. Two FS techniques were employed to classify KOAs, and the suggested FE method was evaluated using various classifiers. The evaluation was conducted using a dataset consisting of 688 X-ray images obtained from the OA project dataset. Experimental results showed that the RF classifier, when combined with the Tamura and Gabor factors, exhibited outstanding performance in KOA diagnosis. A study [18] on human disease prediction utilized Ant-Lion Optimization (ALO) to identify pertinent characteristics associated with four primary ailments: diabetes, diabetic retinopathy, cancer, and heart disease. The strategy successfully reduced the complexity of features while achieving the highest levels of accuracy in datasets related to heart disease, diabetes, diabetic retinopathy, and skin cancer. This study demonstrated the effectiveness of ALO in improving predictive models by significantly reducing the number of features while maintaining performance metrics. Furthermore, a multi-objective approach [19] was employed to minimize test suites in web applications, utilizing the Non-dominated Sorting Genetic Algorithm-II (NSGA-II). The approach effectively balanced between minimizing redundancy and maximizing coverage across multiple software versions, thereby demonstrating the superior performance of NSGA-II compared to other algorithms in optimizing comprehensive test suites. In research [20], the incorporation of the Hybrid Chaotic Flower-Fruit Fly Optimization (Hybrid CFFO) technique for Test Case Prioritization (TCP) led to substantial enhancements in



**Fig. 1.** Proposed methodology workflow.

software testing efficiency. The Hybrid CFFO algorithm effectively prioritized test cases based on metaheuristic principles, outperforming traditional TCP methods in terms of fitness, Normalized.

Average Percentage of Faults Detected, and Average Percentage of Combinatorial Coverage scores. Additionally, a refined method for selecting relevant features in predicting chronic diseases, particularly in breast cancer research, is proposed in study [21]. This method utilizes Teaching Learning-Based Optimization (TLBO), Elephant Herding Optimization (EHO), and a hybrid TLBO-EHO algorithm. This study emphasized the significance of FS in enhancing classification accuracy and performance metrics. These works showcase the evolving methods for FS, highlighting the adaptability and effectiveness of nature-inspired computing techniques in addressing challenging real-world problems across various domains. The Gravitational search optimization algorithm (GSOA) technique is proposed in [22] for metaheuristic-based FS. The initial step is to utilize public and private datasets to obtain 36 features from retinal fundus images of both healthy and glaucoma. Using the subset of features supplied by the GSOA, six ML models are trained for categorization. By focusing on the most important features, the proposed FS method improves classification performance. An effective FS strategy for predicting COVID-19-infected persons from chest CT scans is proposed in this study [23]. Following pre-processing, 213 features were extracted from a public dataset. There is a two-step procedure to determine which features are most important for differentiating between healthy and COVID persons. Several nature-inspired algorithms—a TLBO, a cuckoo search optimization, and a combination of the two—are used to improve the initial 75 % of features retrieved by the Chi-square test. To classify these CT images, five ML classifiers are used. There is a noticeable improvement in accuracy with the proposed FS method.

## 2.2. ML approach

In research [24], the focus is on developing classification models for breast cancer using various ML classifiers. FS is performed using three different approaches. Depending on the best-performing models, the best-performing feature subset is chosen, and an ensemble-based Maximum Voting Classification model is suggested. The suggested algorithm outperforms the competition with an exceptional accuracy score. In the article [25], the research leverages lateral view knee radiographs to develop a model for Patellofemoral Osteoarthritis (PFOA) detection. Automated patellar Region-Of-Interest (ROI) detection is

performed, and three different texture ROIs are extracted, utilizing LBP for feature description. An ML model (Gradient Boosting Machine) and deep CNNs are trained for PFOA detection. Performance is compared with conventional models using clinical assessments and participant characteristics, resulting in classification outcomes based on expert readers' assessments.

## 2.3. DL approach

In the context of KOA classification and localization, study [26] utilizes radiographic images and transforms 2D radiographs into 3D representations. Local Binary Patterns (LBP) and deep features extracted from Alex-Net and Dark-net-53 are used, and the best features are selected using Principal Component Analysis (PCA). These features are then combined and given to classifications for 10-fold cross-validation, yielding a good KOA grade categorization accuracy. The research also provides a localization model that uses an Open Exchange Neural Network (ONNX) and YOLOv2 to locate KOA in categorized images, obtaining an excellent mean Average Precision (mAP), indicating better outcomes when compared to existing methods. In the context of KOA diagnosis from X-ray images, the paper [27] addresses the development of Active Learning (AL) methods. OA is recognized as a substantial societal burden with continually escalating associated costs. Automatic diagnostic methods have the potential to alleviate these costs, and DL techniques hold promise in this regard. While many previous KOA severity grading studies using DL assume access to large annotated datasets for model development, this research showcases the creation of a KOA severity grading model through AL with as few as 50 randomly selected samples from unlabelled data. The significant insight here lies in the enhanced AL performance achieved by incorporating consistency regularization, a common practice in semi-supervised learning.

In research [28], a transfer learning-based model is suggested for distinguishing between the presence and absence of KOA. The model leverages the pre-trained InceptionResNetV2 model to extract deep features from KOA Magnetic Resonance (MR) images sourced from the OAI-ZIB dataset. Two different optimization algorithms, Stochastic Gradient Descent (SGD) and Root Mean Square Propagation (RMSprop), are employed for training and testing the knee MR image dataset. The model's performance is assessed using various evaluation metrics. Experimental results indicate that the transfer learning-based model, particularly when using the RMSprop optimizer, achieves superior classification results in both patient-level and image-level dataset

**Table 1**  
KOA Data samples.

Grade	Actual Count	Finalized Count for Binary classification	Finalized Count for Multi-class classification
0	3857	1000	1000
1	1770	250	1000
2	2578	250	1000
3	1286	250	1000
4	295	250	1000

partitioning methods. This includes achieving high positive metrics, and surpassing the performance of comparison algorithms. These results underscore the model's effectiveness in accurately classifying KOA from MR images. Research [29] proposes the utilization of six pre-trained Deep Neural Network (DNN) algorithms for KOA diagnosis from the Osteoarthritis Initiative (OAI) dataset. They conducted binary classification to determine the presence of KOA and multiclass classification to ascertain the severity of KOA. Testing was performed on three datasets, each comprising different classes of KOA images, for comparison. Utilizing the ResNet101 DNN model, they achieved maximum classification accuracy. The proposed model in the study [30] aims to perform both multi- and binary classifications of KOA severity to ascertain the degree of the condition. It is designed to identify peripheral, disperse, and vascular thickened opacities, aiding doctors in understanding the main causes of KOA. Using the OAI dataset, they evaluated the proposed framework after performing pre-processing steps such as normalization, artifact removal, scaling, and contrast handling. The fine-tuned DenseNet169 model outperformed its predecessors in both multi- and binary classification.

The existing methods for medical images have made significant progress. From the literature, we identified that efficient FS, ML, and DL design are all important for achieving better accuracy. The studies provide better accuracy, but the complexity of the models is very high and requires high-end computers to satisfy their computational requirements. In case the model is simple, accuracy has to be sacrificed. Our study aims to address this difficulty by utilizing an optical FS technique in combination with simple ML models to obtain higher accuracy in grading and detecting KOA. The proposed model is simple, does not require high-end computers, and at the same time, gives excellent accuracy in KOA detection.

### 3. Materials and methods

In this section, we present a detailed account of the sequential

processes employed for the classification of KOA. The section begins with data collection, followed by data processing, FE, FS, and culminating in the implementation of an ML classification model. The entire process of the proposed methodology for KOA prediction is illustrated in Fig. 1.

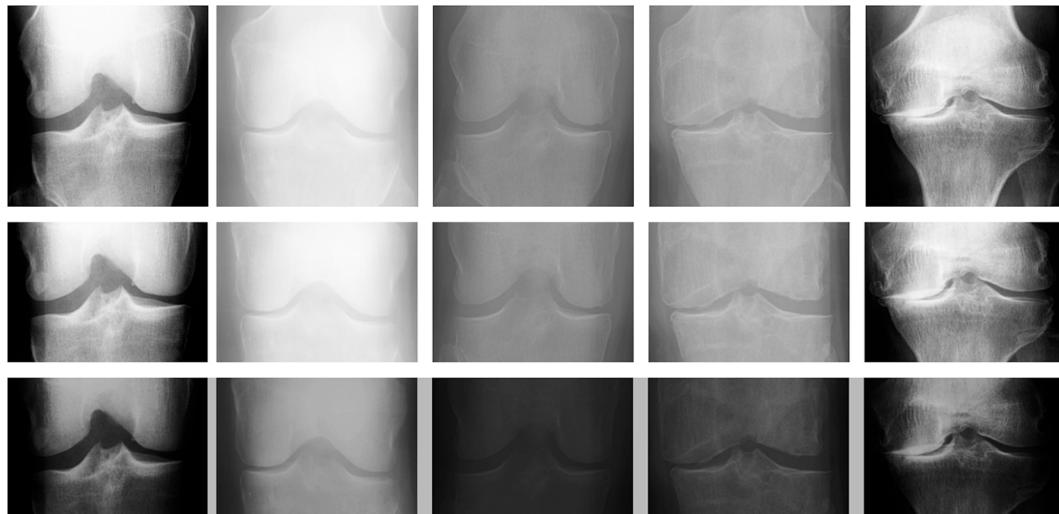
#### 3.1. Data and processing

We obtained our dataset from Kaggle [31], and the dataset can be accessed via the following link: <https://www.kaggle.com/datasets/shashwatwork/knee-osteoarthritis-dataset-with-severity>. This dataset contains images ranging from grade 0 to 4, representing different severity levels of KOA. The grades 0, 1, 2, 3, and 4 correspond to healthy, doubtful, minimal, moderate, and severe KOA, respectively. The dataset contains four folders: auto\_test, train, validate, and test. Combining all the folders, we have 3857, 1770, 2578, 1286, and 295 samples from grades 0 to 4, respectively. To address class imbalance, we employed a data-balancing approach by applying undersampling and oversampling techniques. Specifically, we ensured each category had 1000 samples. The KOA Grading System is a widely accepted grading procedure for its simplicity and reliability. The grading system is based on the Kellgren-Lawrence (KL) classification [32]. The KL classification system assigns a grade from 0 to 4, with higher grades indicating more severe OA. We organized this dataset into two sets: a binary dataset and a multi-dataset. In the binary dataset, grades 0 and 1 are grouped as one category, while grades 2 to 4 are considered another category. The multi-dataset retains the original data structure. The KOA data samples used for binary and multi-classification are detailed in Table 1. 80% of data samples are utilized for training and the rest for testing.

Following dataset organization, we applied image processing techniques to enhance the quality of the images and focus on relevant areas. This involved segmenting the images to remove unwanted regions and applying histogram equalization to improve image quality. Fig. 2 illustrates the results of this process, displaying the original image in the first row, the segmented image in the second row, and the equalized image in the third row.

#### 3.2. Feature extraction

FE from KOA X-ray images using CNN is a critical component of modern medical imaging analysis. CNNs are extremely well-suited for this purpose since they can automatically learn and extract useful data from images [33,34]. The process begins with data collection and pre-processing, where X-ray images are gathered and enhanced to ensure



**Fig. 2.** Original and processed KOA Images with various grading.

clarity and uniformity. A pre-trained VGG architecture [35] is then employed, typically consisting of convolutional layers to detect various patterns and features in the images, followed by pooling and fully connected layers for reducing dimension and learn higher-level features. During the training phase, the CNN learns to recognize specific features associated with KOA, such as variations in bone density, the presence of osteophytes (bone spurs), and changes in joint spaces.

The result of this process is a set of feature maps that highlight different aspects of the X-ray images, capturing both fine and coarse-grained details. These feature maps are then converted into feature vectors, which are numerical representations of the most important features extracted from the images. These feature vectors serve as a condensed and highly informative representation of the original images. It has 200 features. They are invaluable for subsequent tasks, such as classifying the severity of KOA, monitoring disease progression over time, and assisting radiologists in making more accurate diagnoses. Following this extraction process, the GBC and PSO algorithms are employed to identify the most significant features, ultimately narrowing down the feature count to 120. Importantly, both the FE and selection stages are conducted exclusively on the training dataset to mitigate the risk of information leakage, ensuring the integrity of the model evaluation process.

### 3.3. Feature selection

FS indeed plays a crucial role in improving the predictive power of our ML model. We have carefully selected optimized techniques for this purpose, and the intricate workings of these optimization methods are elaborated upon below, accompanied by a flowchart for better understanding. Accuracy is chosen as the fitness function for both algorithms because our primary goal is to enhance the detection rate of KOA prediction. During each iteration of the optimization algorithm, varying accuracies are generated based on the selected features. Upon completion of the iterations, the algorithm finalizes the best features that yield the highest accuracy for further analysis. PSO and GBC algorithms are utilized for FS in medical image classification due to their ability to efficiently explore high-dimensional feature spaces and identify optimal feature subsets. PSO is inspired by the social behavior of birds flocking or fish schooling. It operates by iteratively adjusting the positions of a swarm of particles in the search space based on their individual and collective best-known positions. This iterative process allows PSO to efficiently search for the most discriminative features by optimizing a given evaluation criterion, such as classification accuracy. GBC is inspired by the foraging behavior of bees. Similar to PSO, it operates iteratively to search for optimal solutions. GBC employs a combination of genetic algorithms and the waggle dance behavior of bees to guide the search process. By integrating genetic operators with the exploration-exploitation strategy of bee colonies, GBC can effectively explore the feature space and identify relevant features for classification tasks.

### 3.4. Genetic bee colony (GBC) algorithm

The intelligent swarm technique for identifying and collecting food sources in a bee colony is defined within the framework of GBC [36]. The system uses many factors, like as communication links, task distribution, reproduction, dancing, positioning, mating, and motion, to determine which bee is best suited to solve a given challenge. The GBC algorithm iteratively optimizes the solution to enhance efficiency in addressing critical problems. The bee swarm is divided into three categories: employees, onlookers, and scouts. Employee bees are responsible for discovering new food sources, while scout bees assign fitness scores to locations identified by employee bees for random search. This assignment process is based on randomness. The bees will migrate to a new place to forage if it is determined to have better food than their current one. While onlooker bees evaluate food sources based on quantitative parameters like availability, employee bees are always

looking for the optimum location to collect food [37]. There are four main phases of the GBC method: initialization, employed, onlooker, and scout bee. Most algorithms have an initialization phase accompanied by three primary phases that are run over and over again until the end condition met. The GBC algorithm's main components can be summed up as follows.

#### 3.4.1. Phase 1 (Initialization)

A random distribution of  $SN$  solutions (called food sources) is generated by the GBC method during its initialization phase. In this case,  $SN$  can also stand for the total number of busy bees or curious onlookers. Let's write  $i$  as the problem's magnitude, and the  $i^{th}$  food source as  $\varphi$ . Each food source is created within the defined range of the  $i^{th}$  index using the formula where  $\varphi$  ranges from 1 to 2, up to  $SN$ , and  $j$  ranges from 1 to 2, up to  $D$ .  $\varphi_{ij}$  is a random real number that follows a uniform distribution between a specified range,  $x_{min}$  and  $x_{max}$  being the lower and upper bound for the dimension  $D$ . Furthermore, for each food source, an initialization of a trial counter is performed as detailed in Equation (1).

$$x_{ij} = x_j^{min} + \varphi_{ij}(x_j^{max} - x_j^{min}) \quad [1]$$

#### 3.4.2. Phase 2 (Employed Bee (EB))

In GBC Optimization, EBs are pivotal for optimization. Each EB selects a food source and creates a nearby one. They perform a local search, involving random index selection  $i$  and a distinct food source. A random weight  $\varphi$  is applied to adjust the current food source using Equation (2):

$$v_{ij} = x_{ij} + \varphi(x_{ij} - x_{r1,j}) \quad [2]$$

Here,  $v_{ij}$  is the updated value at position  $(i,j)$ ,  $x_{ij}$  is the current value, and  $x_{r1,j}$  is from a reference source. A greedy selection mechanism picks the better option between  $i$  and  $k$ . This process helps bees explore and refine solutions in their vicinity, contributing to colony optimization.

#### 3.4.3. Phase 3 (Onlooker Bee (OB))

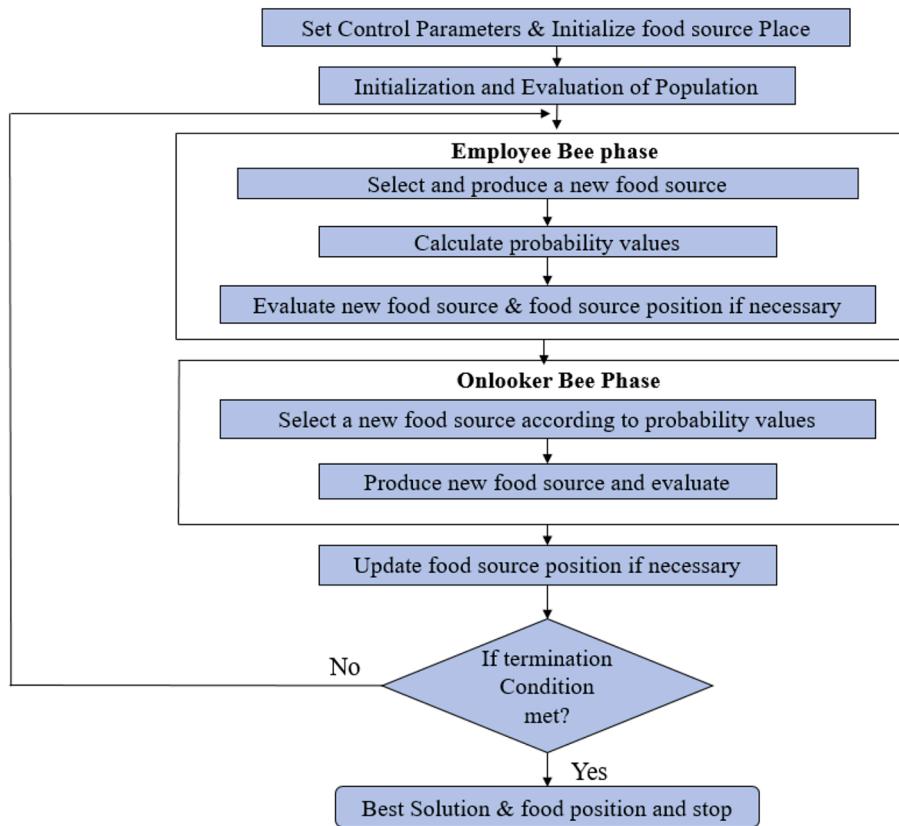
Unlike worker bees, which select food based on proximity, OBs make their selections based on probabilities. The nectar content of each food source is used in Equations (3) and (4) to calculate this probability. These equations consider the fitness value of a solution, in the context of minimization problems. For maximization problems, different fitness functions are applied. OBs employ a roulette wheel-based probabilistic selection mechanism to increase their chances of visiting superior food sources. Consequently, OBs seek to investigate new potential food sources near plausible solutions. An OB creates a new solution once it has chosen a food source. A competitive selection happens among the newly derived and the previous solution, analogous to the EB phase.

$$p_i = \frac{\bar{fit}_i}{\sum_{j=1}^{SN} \bar{fit}_i} \quad [3]$$

$$\bar{fit}_i = \begin{cases} \frac{i}{1 + \bar{fit}_i}, fi \geq 0 \\ 1 + abs(fi), fi < 0 \end{cases} \quad [4]$$

#### 3.4.4. Phase 4 (Scout Bee (SB))

Each food source is associated with a trial counter, indicating the number of attempts made to improve that specific food source. After a specified number of attempts during both the onlooker and EB phases, a food source cannot be enhanced, the EB linked to that food source transitions into the role of a SB. Subsequently, the SB searches for a new food source. The inclusion of the SB phase in the GBC algorithm enables it to effectively escape from local minima and enhance its diversity. It's important to note that although in the EB phase, a local search is performed on each food source, in the OB phase, improvements are made to the food sources. Therefore, the GBC method's EB phase principally



**Fig. 3.** GBC algorithm's workflow.

supports diversification, and the OB phase is accountable for intensification. For effective FS using the GBC algorithm, the hyperparameters include population size, number of iterations, employed bees, onlooker bees, and random weight. Refer to Fig. 3 for the visual representation of the GBC algorithm's flow.

#### 3.4.5. Advantages of GBC

The GBC algorithm offers several merits, making it a valuable optimization technique. It excels in finding global optimization solutions, stands out for its simplicity and flexibility, and requires only a minimal number of control parameters. The GBC algorithm has demonstrated its effectiveness in a wide range of real-world applications. Applications based on the GBC algorithm are known for being easy to implement, robust, fast-converging, adaptable, and time-efficient. Furthermore, when compared to other algorithms, the GBC algorithm involves fewer parameters to consider.

#### 3.4.6. Disadvantages of GBC

Despite its advantages, the GBC algorithm also exhibits certain drawbacks. It tends to have a slower convergence speed when dealing with large-scale computations, and its accuracy may be comparatively lower in such cases. GBC can encounter the issue of premature convergence, particularly in long-duration applications. Another limitation is that the population size remains fixed, and variations and adjustments in population size are not supported. Although allowing individuals to expand the search space can increase the likelihood of discovering global optimization solutions, this approach comes at the cost of consuming significant time in each generation. Conversely, GBC may become trapped in local minima, limiting its ability to find the best solutions.

#### 3.5. Particle swarm optimization (PSO)

PSO approach draws inspiration from the intelligent behaviour of birds [38]. Craig Reynolds [39] was the first to simulate the social flocking behaviour of birds, a concept later explored by Frank Heppner [40]. PSO operates by seeking an optimal solution, much like the co-ordinated flight of birds, with each particle having specific velocities determined based on past outcomes and the behaviours of neighbouring particles in defined search areas. To address a problem within the search space, solutions are represented in various n-dimensional forms, with each of these representations corresponding to a particle in the PSO algorithm. These particles navigate through an n-dimensional solution space at varying velocities. As they move, they store their past behaviours and share their experiences to collectively explore the search space. A key strength of PSO lies in its ability to facilitate information exchange among particles, enabling them to communicate with a portion or the entire swarm to collectively guide their movements in search of the optimal solution. In each iteration, each particle evaluates its current fitness value with prior optimized outcomes as well as the fitness values of its neighbours. Both global and local considerations factor into the algorithm for every particle. The best global value for each particle is preserved as a local value, while the finest outcome among all particles in the complete search space is documented as the global best optimal solution. In subsequent iterations, adjustments are made toward the best optimal solution if the current value surpasses previous results.

In the realm of PSO, each particle operates as an agent with the responsibility of discovering a viable solution within the predefined search space of an optimization problem. The particle's flight behaviour mirrors the exploration process of an individual particle. The particle's velocity undergoes continuous updates, guided by its current and the optimal position of the entire swarm population. This swarm population consists of  $M$  particles. Within this context, the historical optimal

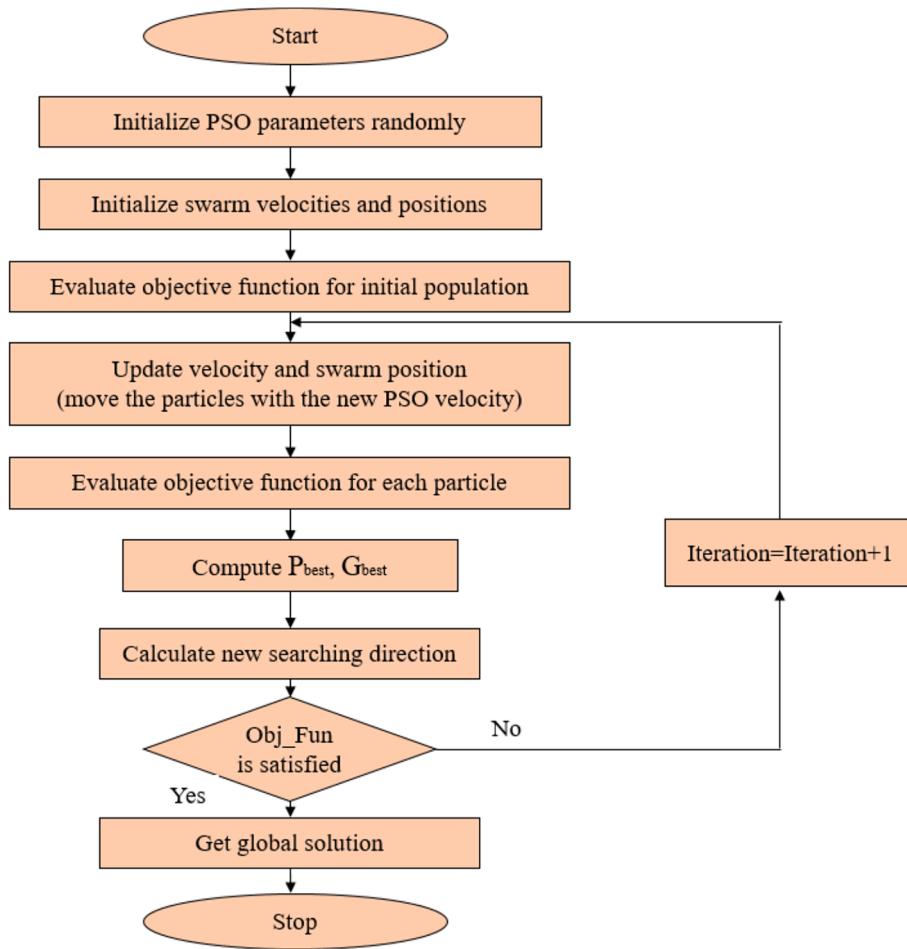


Fig. 4. PSO algorithm workflow.

**Table 2**  
Outcome of Binary Classification without FS [51].

MODEL	SVM	KNN	LDA	RF
ACCURACY	98.85 %	96.25 %	94.98 %	97.64 %
SENSITIVITY	99.35 %	96.48 %	94.92 %	98.07 %
SPECIFICITY	98.20 %	95.95 %	95.07 %	97.09 %
NPV	99.16 %	95.42 %	93.36 %	97.49 %
PPV	98.61 %	96.89 %	96.24 %	97.75 %

position of the  $i^{\text{th}}$  particle is designated as  $p_i$  where  $i$  belongs to the set of natural numbers  $\{1, 2, 3 \dots M\}$ . Additionally, the optimal position for the entire swarm population is denoted as  $g_i$ . Throughout each iteration, both the velocity and position of every individual particle adapt dynamically. These adjustments are made by taking into account each particle's prior positions as well as the optimal position of the entire swarm population. The specific mathematical equations governing these

dynamic adjustments are expressed as follows:

$$V_{ij}^{t+1} = \omega V_{ij}^t + c_1 r_1 (p_{best_{ij}} - X_{ij}^t) + c_2 r_2 (g_{best_j} - X_{ij}^t) \quad [5]$$

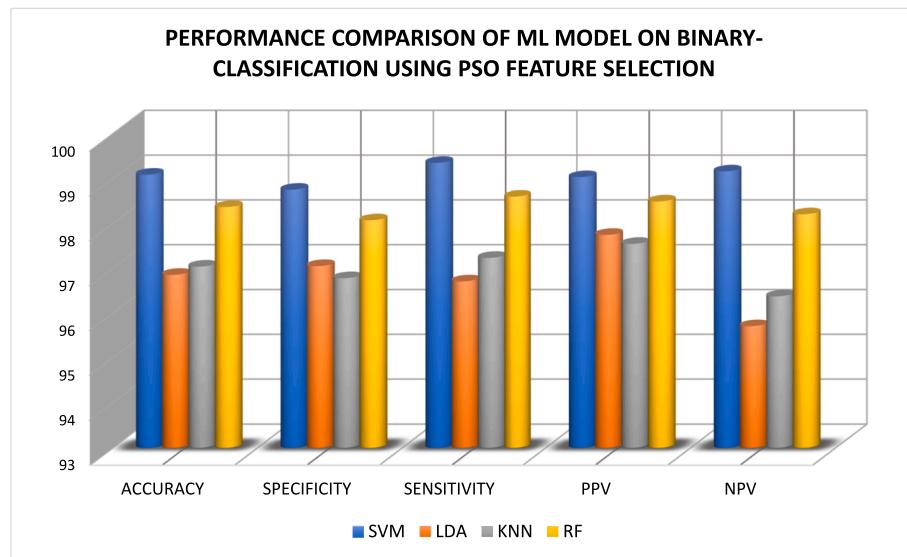
$$X_{ij}^{t+1} = X_{ij}^t + V_{ij}^{t+1} \quad [6]$$

Where  $i = 1, 2, \dots, P$  and  $j = 1, 2, \dots, n$ .

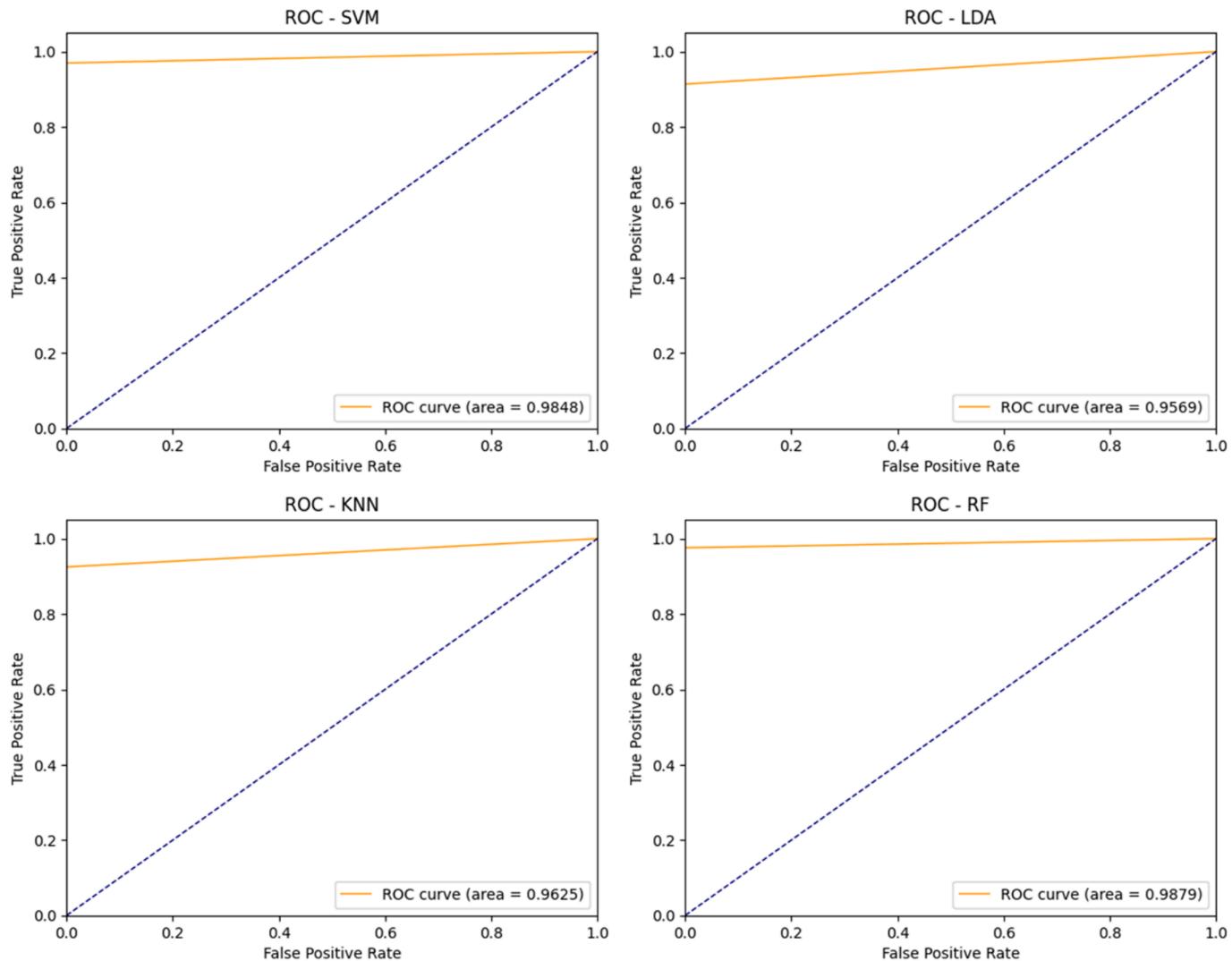
Equation (5) represents the presence of three distinct components contributing to a particle's movement during an iteration. Consequently, Equation (5) comprises three terms, each of which warrants further discussion. Conversely, Equation (6) pertains to the update of a particle's positions. Within this context, the parameter  $\omega$  holds significance as it serves as the inertia weight constant. The significance of this parameter lies in its role in balancing the global and local search, commonly referred as exploration and exploitation. An important observation concerning this parameter is that it represents one of the primary distinctions between the classical PSO version and its derivative

**Table 3**  
Outcome and improvement of Binary Classification using PSO.

MODEL	SVM		KNN		LDA		RF	
	Actual	Improvement	Actual	Improvement	Actual	Improvement	Actual	Improvement
ACCURACY	99.09 %	0.24 %	97.04 %	0.78 %	96.85 %	1.87 %	98.36 %	0.72 %
SENSITIVITY	99.35 %	0 %	97.23 %	0.75 %	96.71 %	1.78 %	98.60 %	0.52 %
SPECIFICITY	98.75 %	0.54 %	96.78 %	0.82 %	97.05 %	1.98 %	98.07 %	0.98 %
NPV	99.16 %	0 %	96.37 %	0.95 %	95.71 %	2.35 %	98.20 %	0.71 %
PPV	99.03 %	0.42 %	97.54 %	0.64 %	97.74 %	1.50 %	98.49 %	0.73 %
F-Measure	99.19 %	—	97.39 %	—	97.22 %	—	98.54 %	—



**Fig. 5.** ML Model comparison on binary classification using PSO's features.

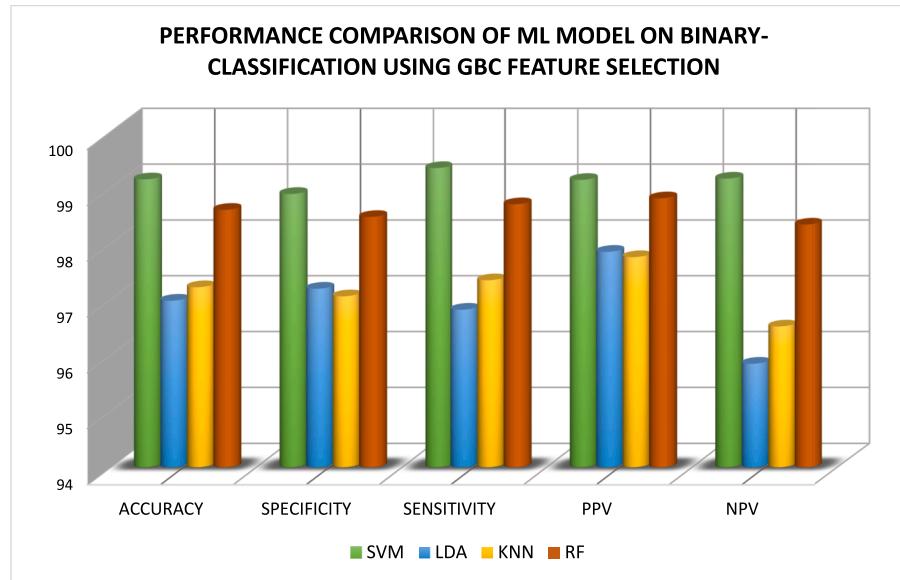


**Fig. 6.** AUC-ROC plot for binary classification using PSO features.

**Table 4**

Outcome and improvement of Binary Classification using GBC.

MODEL	SVM		KNN		LDA		RF	
	Actual	Improvement	Actual	Improvement	Actual	Improvement	Actual	Improvement
ACCURACY	99.15 %	0.30 %	97.22 %	0.96 %	96.98 %	1.99 %	98.61 %	0.96 %
SENSITIVITY	99.35 %	0 %	97.34 %	0.86 %	96.81 %	1.89 %	98.71 %	0.63 %
SPECIFICITY	98.89 %	0.68 %	97.05 %	1.10 %	97.19 %	2.12 %	98.48 %	1.39 %
NPV	99.16 %	0 %	96.51 %	1.09 %	95.85 %	2.48 %	98.34 %	0.85 %
PPV	99.14 %	0.53 %	97.76 %	0.86 %	97.85 %	1.60 %	98.81 %	1.05 %
F-Measure	99.25 %	—	97.55 %	—	97.33 %	—	98.76 %	—

**Fig. 7.** ML Model comparison on binary classification using GBC's features.

versions.

The initial component in the velocity update equation corresponds to the result of multiplying the parameter  $\omega$  by the particle's past velocity, signifying the impact of a particle's preceding movement on its current path. For instance, when  $\omega = 1$ , the particle's motion is entirely determined by its preceding movement, resulting in a continuation of its previous direction. Conversely, when  $0 \leq \omega \leq 1$ , this influence diminishes, leading the particle to explore other regions within the search domain. Consequently, a reduction in the inertia weight parameter promotes broader exploration of the search domain by the swarm, potentially increasing the likelihood of discovering a global optimum. However, employing lower values of  $\omega$  comes at a cost, namely, an increase in the computational time required for simulations.

The second term in Equation (5), referred to as individual cognition, is calculated as the difference  $(pbest_{ij} - X_{ij}^t)$  between the particle's best ( $pbest_{ij}$ ) and its current position ( $X_{ij}^t$ ). This term's purpose is to increase as the particle moves farther away from its best position, thereby attracting the particle back toward its optimal location. The parameter  $c_1$  is a constant. It determines the significance of the particle's past experiences in influencing its movement. The other factor in this term is  $r_1$ , a random value parameter with a range of [0,1]. The random parameter takes a crucial role in preventing premature convergence and increasing the likelihood of finding the global optimum.

The third element in Equation (5) is social learning, which allows all particles to share knowledge on the best point attained ( $gbest_i$ ), independent of whose particle discovered it first. Its structure resembles that of the second component associated with individual learning. Consequently, akin to the individual learning component, the disparity  $(gbest_{ij} - X_{ij}^t)$  serves as an attractive force guiding particles towards the

best point detected during the current iteration. Analogous to  $c_1$ , the parameter  $c_2$  embodies a social learning parameter, governing the significance of global learning within the swarm. Similarly,  $r_2$  functions in the same manner as  $r_1$ , introducing randomness into the process to forestall premature convergence.

Finally, Fig. 4 presents the flowchart of the PSO algorithm. Notably, this algorithm is designed to search for minima, with all position vectors being evaluated by the fitness function  $f(X)$ . When configuring PSO for optimization tasks, key parameters such as swarm size, inertia weight, acceleration and iteration are specified to ensure effective performance. The swarm size influences the algorithm's exploration capabilities, with larger swarms facilitating broader exploration but at the expense of increased computational complexity. Inertia weight ( $w$ ) dictates the balance between exploration and exploitation. Cognitive ( $c_1$ ) and social ( $c_2$ ) acceleration coefficients regulate a particle's movement towards personal and global best positions, respectively. The maximum iterations determines the algorithm's termination, ensuring sufficient time for convergence without excessive computational burden. Through careful tuning of these parameters, the PSO efficiently navigate solution spaces to find optimal solutions for various optimization tasks.

### 3.5.1. Advantages of PSO

- PSO exhibits robust communication capabilities among particles, facilitated by the storage of the best particle positions from previous iterations.
- PSO accelerates towards optimal solutions in optimization problems, promoting faster convergence.
- Simplicity: PSO boasts straightforward equations for velocity and position updates, making calculations relatively easy.

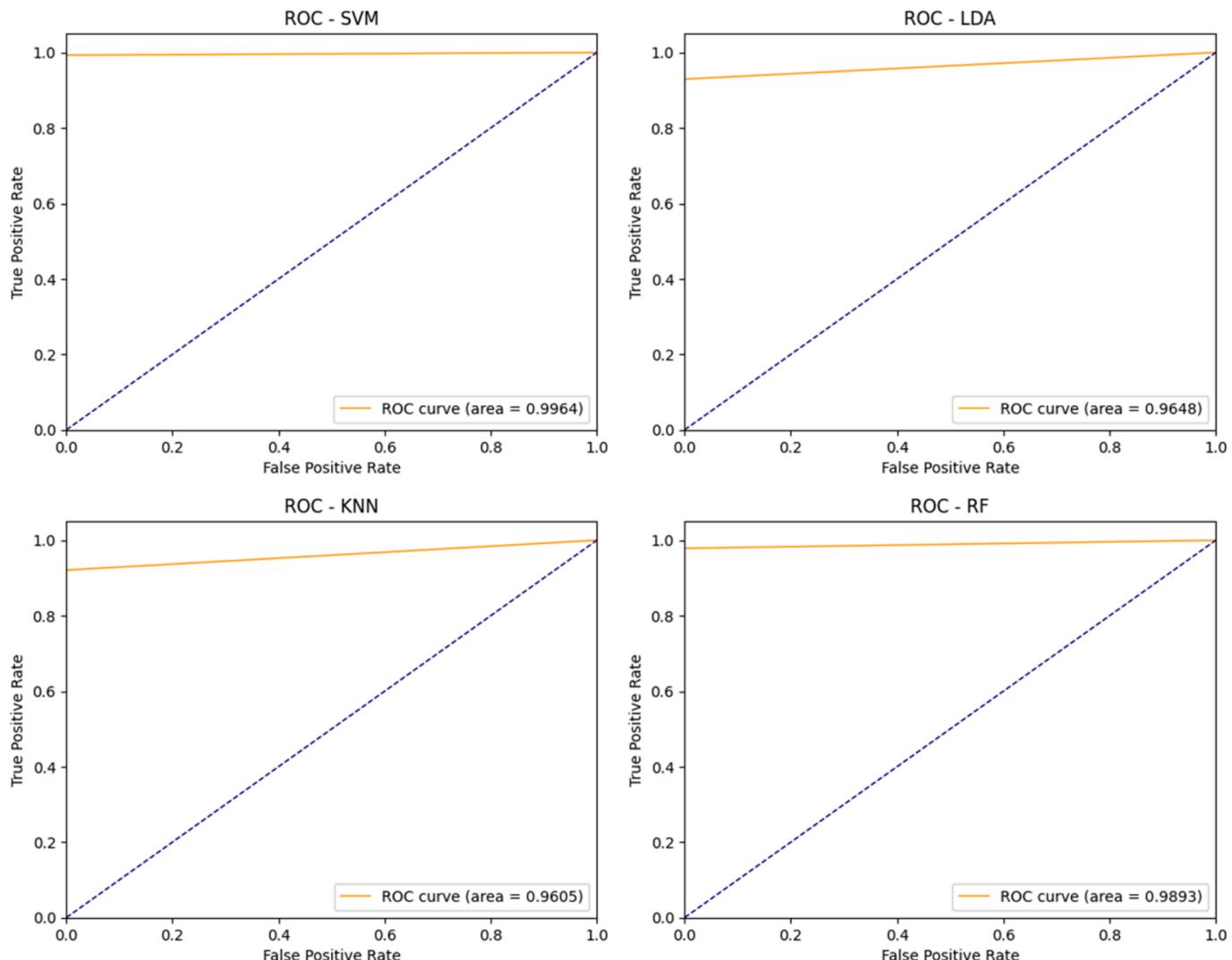


Fig. 8. AUC-ROC plot for binary classification using GBC features.

**Table 5**  
Outcome of Multi Classification without FS [51].

MODEL	SVM	KNN	LDA	RF
ACCURACY	95.59 %	92.93 %	90.88 %	96.37 %
SENSITIVITY	97.25 %	95.00 %	93.10 %	98.02 %
SPECIFICITY	93.54 %	90.46 %	88.14 %	94.36 %
NPV	96.53 %	93.81 %	91.21 %	97.50 %
PPV	94.86 %	92.24 %	90.62 %	95.50 %

- PSO demonstrates the ability to adapt to changing environments by selecting optimal solutions that suit evolving conditions.

### 3.5.2. Disadvantages of PSO

- PSO assumes all particles are identical, which can limit its effectiveness in scenarios where particles should exhibit diverse behaviours.
- Single Optima Focus: PSO does not inherently identify multiple optima in a solution space, potentially missing valuable solutions.
- Achieving convergence can be challenging when dealing with varying inertia weights, adding complexity to optimization processes.

**Table 6**  
Outcome and improvement of Multi Classification using PSO.

MODEL	SVM		KNN		LDA		RF	
	Actual	Improvement	Actual	Improvement	Actual	Improvement	Actual	Improvement
ACCURACY	97.28 %	1.69 %	95.53 %	2.59 %	94.68 %	3.80 %	98.73 %	2.35 %
SENSITIVITY	97.91 %	0.66 %	96.39 %	1.39 %	94.48 %	1.37 %	98.70 %	0.67 %
SPECIFICITY	96.50 %	2.95 %	94.45 %	3.99 %	94.94 %	6.80 %	98.77 %	4.40 %
NPV	97.41 %	0.88 %	95.49 %	1.67 %	93.16 %	1.95 %	98.36 %	0.86 %
PPV	97.17 %	2.30 %	95.56 %	3.32 %	95.93 %	5.30 %	99.02 %	3.51 %
F-Measure	97.54 %	—	95.97 %	—	95.20 %	—	98.86 %	—

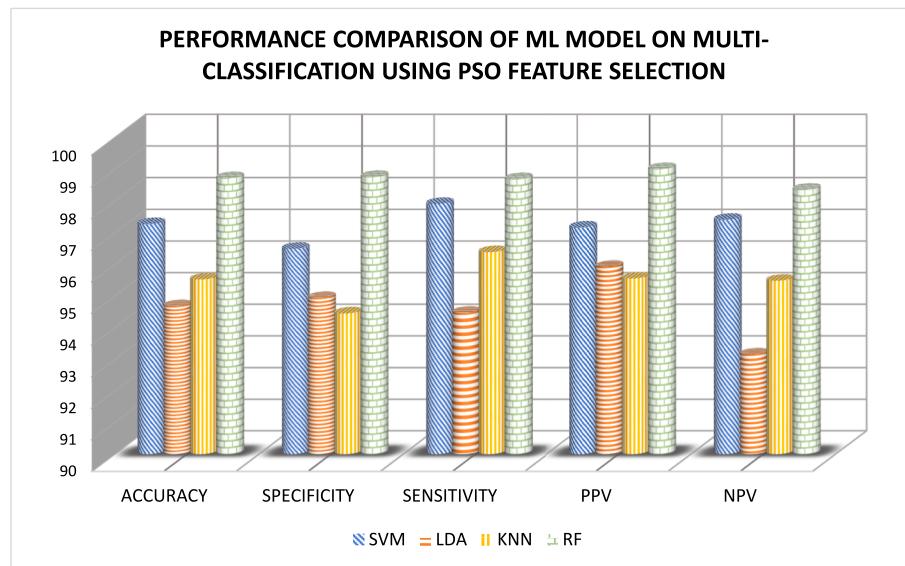


Fig. 9. ML Model comparison on multi classification using PSO's features.

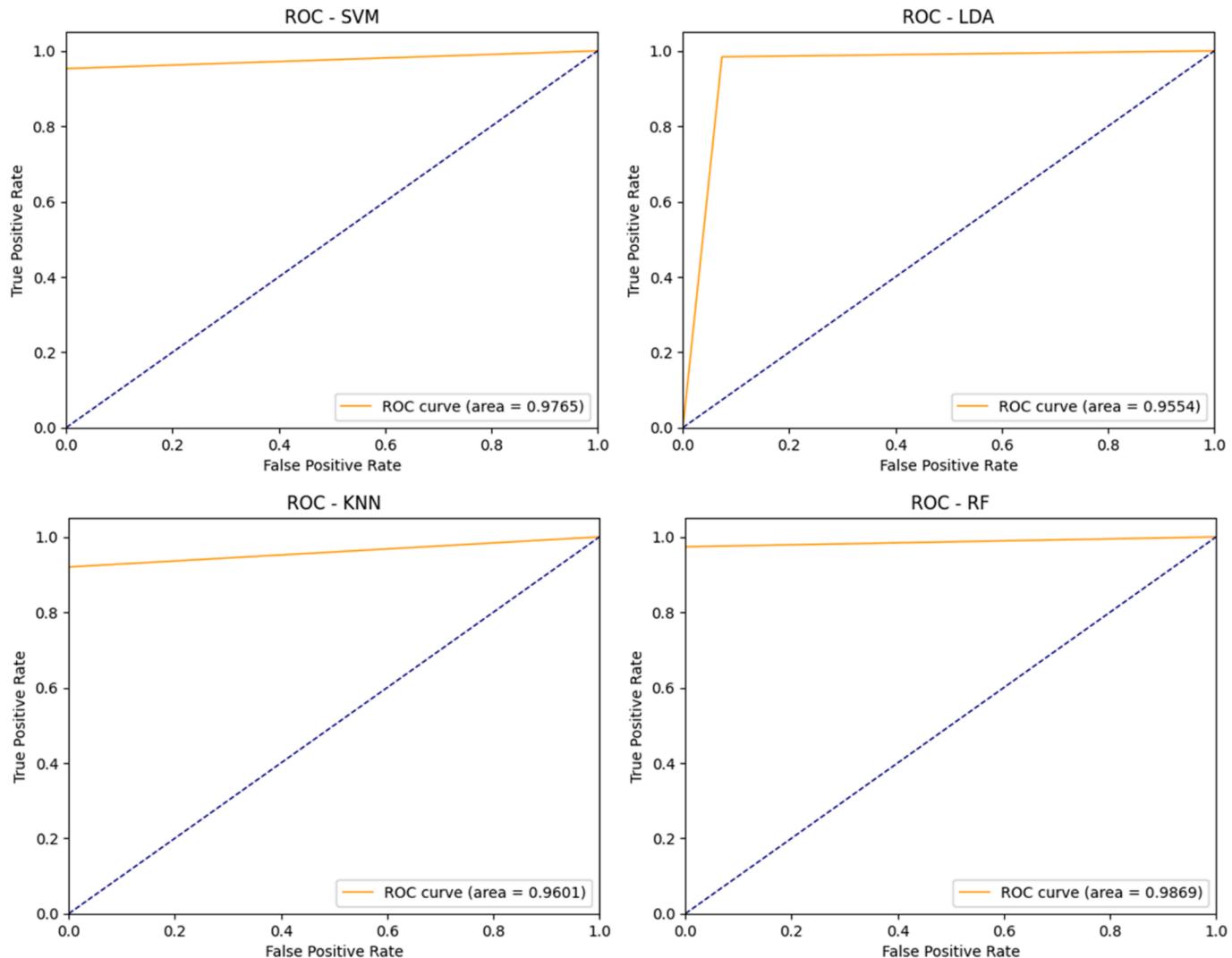
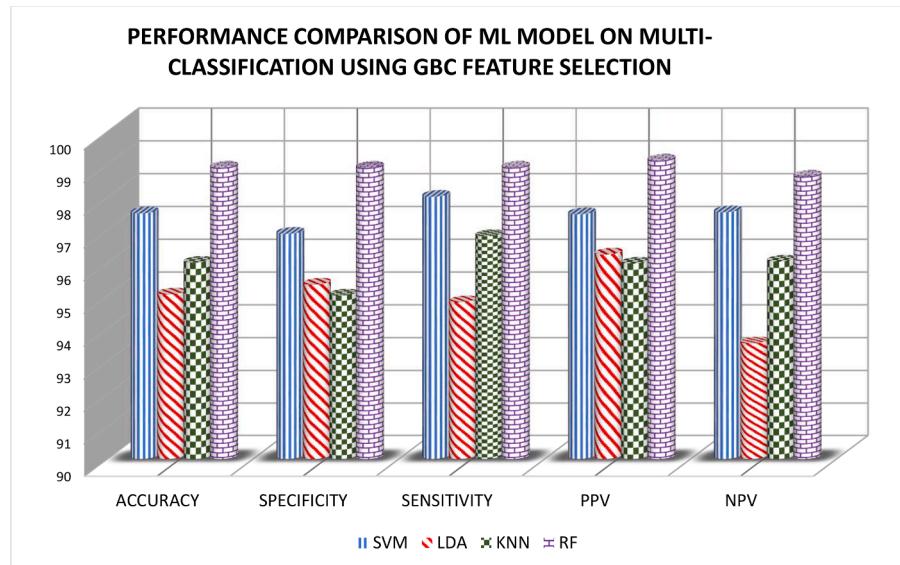


Fig. 10. AUC-ROC plot for binary classification using PSO features.

**Table 7**

Outcome and improvement of Multi Classification using GBC.

MODEL	SVM		KNN		LDA		RF	
	Actual	Improvement	Actual	Improvement	Actual	Improvement	Actual	Improvement
ACCURACY	97.52 %	1.93 %	96.01 %	3.07 %	95.04 %	4.16 %	98.91 %	2.53 %
SENSITIVITY	98.03 %	0.77 %	96.82 %	1.82 %	94.81 %	1.70 %	98.91 %	0.89 %
SPECIFICITY	96.90 %	3.35 %	95.01 %	4.54 %	95.34 %	7.20 %	98.90 %	4.54 %
NPV	97.55 %	1.02 %	96.04 %	2.23 %	93.55 %	2.34 %	98.63 %	1.13 %
PPV	97.49 %	2.63 %	95.98 %	3.74 %	96.26 %	5.63 %	99.13 %	3.62 %
F-Measure	97.76 %	1.19 %	96.40 %	2.45 %	95.53 %	2.61 %	99.02 %	1.43 %

**Fig. 11.** ML Model comparison on multi classification using GBC's features.

### 3.6. Classification

In this section, we provide a comprehensive explanation of the ML model that we employed for KOA classification.

#### 3.6.1. Support vector machine

SVM represents a relatively new ML method rooted in statistical learning theory and pioneered by Vapnik [41]. SVM is categorized as a computational approach that follows the principle of structural risk minimization (SRM). This principle allows SVM to derive decision-making rules that yield minimal errors on independent test sets, making it an efficient solution for learning problems. SVM has gained popularity for its ability to tackle various challenges such as nonlinearity, local minima, and high-dimensionality in practical applications. In various real-world situations, SVM consistently demonstrates superior accuracy in long-term predictions when compared to alternative computational methods. SVM functions by utilizing the concept of decision planes, which establish boundaries between distinct classes [42]. By employing a linear model and transforming input vectors through nonlinear mappings into high-dimensional feature spaces, SVM can handle nonlinear class boundaries effectively. SVM addresses complex, nonlinear dependencies [43], such as those present in the mapping function  $y = f(x)$ , connecting high-dimensional input vectors  $x$  to scalar output  $y$  (or vector output  $y$ , as in the case of multiclass SVM).

#### 3.6.2. k-nearest neighbour

The kNN algorithm is a case-based learning technique that preserves all training data for classification tasks. While this approach has its merits, it is considered “lazy” learning [44], which can be limiting in certain applications, particularly when dealing with dynamic web

mining involving large data repositories. To enhance its efficiency, one strategy involves selecting representative instances to stand in for the entire training dataset. This process entails constructing an inductive learning model from the training sample and employ this model (represented by the selected instances) for classification. Numerous algorithms, including Decision Tree (DT) and neural networks, have been developed for this purpose. The performance of these algorithms is typically evaluated based on various criteria. Despite its simplicity, kNN remains a highly effective classification method and is often considered one of the most efficient ones.

In kNN, the algorithm determines the class of a query example by considering the KNN from the training dataset. To find these nearest neighbours, the algorithm calculates the Euclidean distance between the query example and all training instances [45]. The Euclidean distance [46], denoted as  $dt$  and computed as  $dt = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ , provides a measure of distance between the query point  $(x_1, y_1)$  and each data point  $(x_2, y_2)$ .

### 3.7. Random forest

In 2001, Breiman [47] introduced an integrated learning model called RF, which is based on DTs as its fundamental classifiers. The RF algorithm employs the bootstrap method to obtain multiple subsets of samples [48]. It creates a DT using each of these sample subsets and combines the results from several DTs into an RF. When classifying a given sample, the final classification outcome is determined through a majority vote among the DTs. In the classification process of the RF algorithm, where each base classifier's classification results collectively form a distribution of errors, the final refinement of the classification result is achieved. It processes the test features by applying the rules

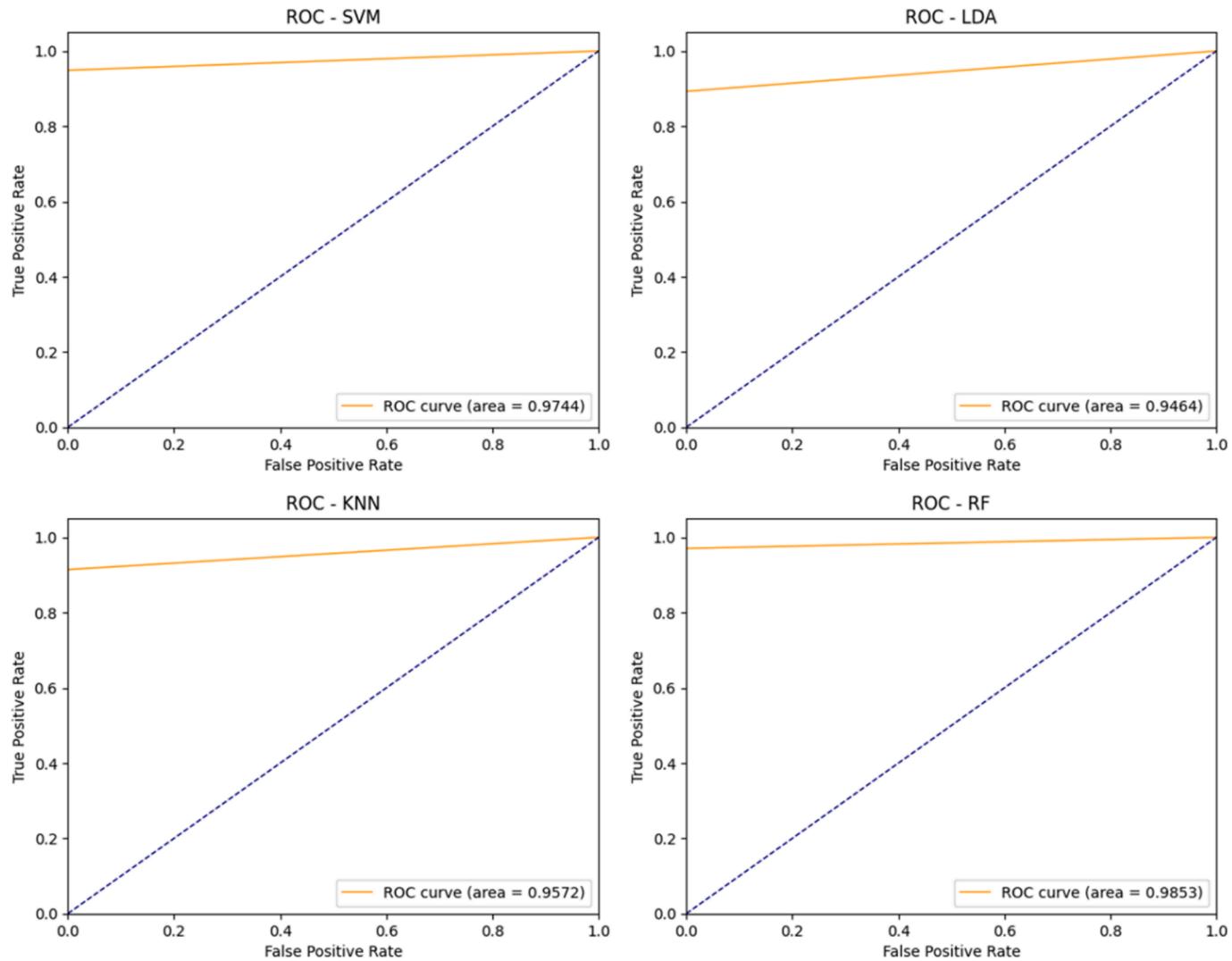


Fig. 12. AUC-ROC plot for binary classification using GBC features.

Table 8

Comparative analysis between the proposed method and state-of-the-art approaches.

REF	MODEL	ACCURACY	SPECIFICITY	SENSITIVITY	PPV
[52]	BiLSTM	84.09 %	—	99.11 %	92.5 %
[30]	DenseNet169	95.93 %	95.41 %	88.77 %	85.8 %
[53]	RefneDet	—	99.36 %	97.54 %	87.22 %
[54]	CNN	84 %	76 %	90 %	73 %
[55]	Ensemble CNN	87 %	87 %	—	86 %
[29]	ResNet101	89 %		86 %	87 %
[56]	CNN	91.03 %		90.4 %	89 %
Ours	GBC+RF	98.91 %	98.90 %	98.91 %	99.13 %

generated by multiple randomly created DTs to make predictions. These predictions, also known as targets, are stored for each DT. The algorithm then tallies the votes for each predicted outcome, and the final prediction made by the RF algorithm is the one with the highest number of votes. The RF algorithm involves two main steps: RF formation and making predictions using the RF classifier. A subset of “K” features is randomly chosen from a larger set of “m” features, with “K” much smaller than “m.” The algorithm then identifies optimal split points among these “K” features, resulting in the creation of new nodes. This

process iteratively continues until a specified number of nodes, “n,” is reached. To form a robust ensemble, this entire procedure is repeated “n” times, ultimately producing a forest of “n” trees. This approach forms the core of the algorithm, efficiently constructing decision trees for predictive modelling.

After generating the RF classifier in the next step, predictions are made. The pseudo-code outlining the prediction process in the RF algorithm involves utilizing the test features to apply the rules generated by each randomly created DT. This application of rules yields predictions, which are then stored as target results. Subsequently, the algorithm computes the vote count for each predicted outcome. The final prediction made by the RF algorithm is determined by selecting the predicted outcome with the highest number of votes. This process is mathematically represented by the equation:

$$H(x) = \operatorname{argmax}_Y \sum_{i=1}^k I(h_i(x) = Y) \quad [7]$$

Where  $x$  is the test sample,  $h_i$  indicates individual DT,  $Y$  stands for output variable (i.e., classification label),  $I$  is the indicator function and  $H$  is the RF model.

### 3.8. Linear discriminant analysis

LDA serves as both a dimensionality reduction and classification

technique with the primary goal of discovering a linear combination of attributes that effectively distinguishes two or more classes [49]. It is particularly beneficial in scenarios involving labelled data, where the aim is to condense dimensionality while preserving the ability to differentiate between classes [50]. Here's a breakdown of the LDA approach for classification, accompanied by the relevant formulae: Assume we possess a labeled dataset comprising  $N$  instances, with the data distributed across  $K$  distinct classes. We initiate the process by calculating the mean vector for each class, representing the average feature values within each class. Mean vector calculation for class  $k$ :

$$\mathbf{m}_k = \frac{1}{N_k} \sum_{i=1}^N \mathbf{x}_i \quad [8]$$

Where  $\mathbf{m}_k$  signifies the mean vector for class  $k$ ,  $N_k$  indicates the number of samples in class  $k$  and  $\mathbf{x}_i$  denotes a sample within class  $k$ . Next, we assess the within-class scatter matrix ( $S_w$ ), which quantifies data dispersion within each class, and the between-class scatter matrix ( $S_B$ ), which gauges the separation of class means relative to one another.

Within-class scatter matrix ( $S_w$ ):

$$S_w = \sum_{k=1}^K \sum_{i=1}^{N_k} (\mathbf{x}_i - \mathbf{m}_k)(\mathbf{x}_i - \mathbf{m}_k)^T \quad [9]$$

Between-class scatter matrix ( $S_B$ ):

$$S_B = \sum_{k=1}^K N_k (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})^T \quad [10]$$

Where  $\mathbf{m}$  represents the overall mean across all samples,  $\mathbf{m}_k$  is the mean of class  $k$ , and  $N_k$  denotes the number of samples in class  $k$ . The subsequent step involves solving a generalized eigenvalue problem to ascertain the linear discriminants (eigenvectors) and their corresponding eigenvalues:

$$S_w^{-1} S_B \mathbf{v} = \lambda \mathbf{v} \quad [11]$$

Where  $\mathbf{v}$  represents the eigenvector and  $\lambda$  signifies the eigenvalue. Following this, we select discriminant directions by arranging the eigenvalues in descending order and opting for the top  $L$  eigenvectors (discriminant directions) corresponding to the  $L$  largest eigenvalues. These selected discriminant directions constitute a novel feature subspace. The original data is then projected onto these  $L$  chosen discriminant directions to obtain a fresh set of features:

$$\mathbf{y} = \mathbf{X} \cdot \mathbf{W} \quad [12]$$

Where  $\mathbf{y}$  represents the transformed data,  $\mathbf{X}$  denotes the original data matrix and  $\mathbf{W}$  is the matrix comprising the top  $L$  eigenvectors. Lastly, after the data transformation, classification can be executed employing a classifier such as logistic regression, SVMs, or any other suitable classification algorithm, utilizing the reduced feature set  $\mathbf{y}$  for decision-making.

#### 4. Results and discussions

In this section, we present a thorough and inclusive examination of the outcomes derived from our ML model. Our analysis encompasses two critical aspects: binary classification and multi-classification scenarios. We explore the impact of employing the FS process, shedding light on the performance variations observed in each case. By comparing results with and without FS, we aim to provide a holistic view of how this critical step influences the model's predictive accuracy and effectiveness.

##### 4.1. KOA binary classification

In the pursuit of optimized FS to improve accuracy in the severity classification of KOA using ML, we conducted a comparative analysis of

two FS methods against the outcomes of previous research [51]. The performance metrics on the binary dataset from the previous study are summarized in Table 2. Table 3 presents the performance results of our ML model using the PSO method, accompanied by the percentage improvement when compared to the previous research. For SVM, we observed an accuracy increase of 0.24 %, while LDA showed a substantial accuracy improvement of 1.87 %. Additionally, KNN exhibited an accuracy boost of 0.78 %, and RF demonstrated an accuracy enhancement of 0.72 %. Further analysis reveals that the PSO method led to improvements in specificity, sensitivity, PPV, and NPV across these ML models. Fig. 5 demonstrate the performance of different ML models when utilizing features selected by PSO for binary classification. Among these models, SVM exhibits the highest accuracy, sensitivity, and NPV, while RF also demonstrates strong performance across all metrics. Fig. 6 illustrates the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) (AUC-ROC) measure plot for binary classification using PSO features across all ML models.

Table 4 provides a comprehensive overview of the performance results achieved by our ML model using the GBC method, along with the percentage improvement when compared to previous research. By employing the GBC method for FS, the accuracy will increase by 0.30 % for SVM, 1.99 % for LDA, 0.96 % for KNN and RF. Next, specificity will increase by 0.68 % for SVM, 2.12 % for LDA, 1.1 % for KNN and 1.3 % for RF. Then, the sensitivity will not increase for SVM, 1.89 % for LDA, 0.86 % for KNN, and 0.63 % for RF. Next, PPV will increase by 0.53 % for SVM, 1.6 % for LDA, 0.86 % for KNN and 1.05 % for RF. Next, NPV will not increase for SVM, 2.4 % for LDA, 1.09 % for KNN, and 0.85 % for RF. Moreover, Fig. 7 presents a bar plot comparing ML models for binary classification using features selected by the GBC algorithm. This bar plot illustrates the performance of different ML models when utilizing features selected by the GBC algorithm for binary classification. Similar to the results obtained using PSO's features, SVM achieves the highest metrics, followed closely by RF. Fig. 8 illustrates the AUC-ROC measure plot for binary classification using GBC features across all ML models.

##### 4.2. KOA severity grading

Table 5 shows previous research's [51] multi-class classification ML model performance. Table 6 presents our ML model results with PSO, including percentage improvements compared to the prior study. By using PSO for FS, we observed accuracy boosts: SVM by 1.69 %, LDA by 3.8 %, KNN by 2.5 %, and RF by 2.35 %. Specificity increased by 2.95 % for SVM, 6.8 % for LDA, 3.99 % for KNN, and 4.4 % for RF. Sensitivity improved by 0.66 % for SVM, 1.37 % for LDA, 1.39 % for KNN, and 0.67 % for RF. PPV increased by 2.3 % for SVM, 5.3 % for LDA, 3.32 % for KNN, and 3.51 % for RF. NPV rose by 0.88 % for SVM, 1.95 % for LDA, 1.67 % for KNN, and 0.86 % for RF. Fig. 9 visually illustrates the ML model's multi-class classification performance using PSO-selected features. This comparison illustrates the performance of different ML models when utilizing features selected by PSO for multi-class classification. RF achieves the highest metrics among the models evaluated, highlighting its effectiveness for multi-class classification tasks. Fig. 10 illustrates the AUC-ROC measure plot for multi classification using PSO features across all ML models.

Table 7 shows the performance of our ML model using the GBC method for multi-classification and the percentage improvements compared to prior research. With the GBC, we observed significant accuracy enhancements: SVM by 1.93 %, LDA by 4.1 %, KNN by 3.07 %, and RF by 2.53 %. Specificity increased by 3.35 % for SVM, 7.2 % for LDA, and 4.54 % for KNN, and RF. Sensitivity improved by 0.77 % for SVM, 1.7 % for LDA, 1.82 % for KNN, and 0.89 % for RF. PPV increased by 2.63 % for SVM, 5.63 % for LDA, 3.74 % for KNN, and 3.62 % for RF. NPV rose by 1.02 % for SVM, 2.34 % for LDA, 2.23 % for KNN, and 1.13 % for RF. Fig. 11 visually illustrates the performance of multi-class classification using GBC features. This comparison highlights the performance of different ML models when utilizing features selected by the

GBC algorithm for multi-class classification. RF stands out as the top performer across all metrics, showcasing its effectiveness in handling multi-class classification tasks. Fig. 12 illustrates the AUC-ROC measure plot for multi classification using GBC features across all ML models.

Our findings represent a significant advancement in the field of KOA detection when benchmarked against current state-of-the-art studies. Across various metrics, our model demonstrates remarkable superiority. In comparison to established models such as BiLSTM [52], DenseNet169 [30], RefneDet [53], CNN [54], Ensemble CNN [55], ResNet101 [29], and another CNN [56], our approach consistently outperforms them all. Notably, our model achieves an accuracy of 98.91 %, specificity of 98.91 %, sensitivity of 98.92 %, and PPV of 99.13 %. Table 8 clearly give previous research and method used and their performance metrics. These exceptional results underscore the efficacy and reliability of our methodology, signifying a substantial leap forward in KOA detection capabilities.

#### 4.3. Discussion

The results of our study provide valuable insights into the effectiveness of FS methods and ML models for KOA detection and severity classification. We investigated the performance of two FS techniques, namely PSO and GBC, in optimizing ML models for this purpose. Our findings indicate that both PSO and GBC methods significantly enhance the performance of ML models when compared to previous research efforts. In our comparative analysis, SVM and RF emerged as top-performing models across binary and multi-class classification tasks. These models consistently demonstrated superior accuracy and other performance metrics compared to LDA and KNN. Our study showcased the robustness and reliability of PSO and GBC methods in identifying relevant features from medical imaging data. These techniques effectively balanced exploration and exploitation, leading to improved model performance and reliability. Notably, RF consistently outperformed other ML models, highlighting its versatility and suitability for KOA detection and severity classification. By benchmarking our results against current state-of-the-art studies, we demonstrated significant advancements in KOA detection capabilities. Our optimized ML models achieved remarkable performance, surpassing the performance of established models such as BiLSTM, DenseNet169, RefneDet, and others. These findings underscore the efficacy and reliability of our methodology, indicating its potential to enhance patient care and treatment outcomes in orthopedic medicine.

Several assumptions are made in the research. The data are not collected directly from a health center but from the Kaggle website. We assumed that the data provided by Kaggle are accurate and that the classification grading system is also correct. We made the initial parameters of the FS algorithm randomly, assuming this approach would yield better results. The proposed approach was tested only on the Kaggle dataset; validation on other datasets and real-time data was not conducted. Lack of validation may restrict the applicability of the proposed approach in real-time healthcare system.

#### 5. Conclusion

In this research, we created a model that can accurately determine the severity of KOA from X-ray scans, and it has proven to be both robust and reliable. The primary objective of our research was to emphasize the crucial role that features play in identifying the severity of KOA disease. To demonstrate this, we utilized data sourced from Kaggle, systematically processed the X-ray images, and extracted a comprehensive set of features. The optimal selected features, and the features extracted using CNN, were subsequently input into our ML model. We rigorously compared the outcomes, and our results conclusively demonstrated that the ML model's correct prediction rate substantially increased when optimization was applied to the features. This research has effectively highlighted the paramount significance of FS in the context of KOA

detection and severity classification using ML models. While our study provides valuable insights and contributions to the field, it's important to acknowledge its limitations for a comprehensive understanding. Firstly, the scope of our research may be constrained by factors such as dataset size, diversity, and quality. Limited access to comprehensive datasets may impact the generalizability of our findings and the robustness of our models. Additionally, the performance of our models may be influenced by various external factors such as variations in imaging protocols, equipment, and patient demographics, which could introduce biases and affect the reproducibility of results. Moreover, while we strive to optimize our methodologies and techniques, there may exist alternative approaches or algorithms that could yield different results or further enhance the performance of our models. Finally, the clinical applicability and real-world efficacy of our findings may require validation through additional studies, including prospective clinical trials or deployment in real-world healthcare settings. Addressing these limitations through continued research, collaboration, and validation efforts will be essential for advancing the field and translating our findings into meaningful clinical practice. Expanding the scope of our research to encompass multiple joint arthritis detection would be valuable for comprehensive orthopaedic diagnostics. In conclusion, the pursuit of improved and comprehensive arthritis detection models is an exciting prospect for future research, with the potential to revolutionize the diagnosis and management of musculoskeletal conditions, ultimately benefiting patient outcomes and healthcare practices.

#### CRediT authorship contribution statement

**Anandh Sam Chandra Bose:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Srinivasan C:** Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Immaculate Joy S:** Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### References

- [1] D.J. McCormack, D. Puttock, S.P. Godsiff, Medial compartment osteoarthritis of the knee: a review of surgical options, *EFORT Open Reviews* 6 (2) (2021) 113–117.
- [2] P.C. Birchfield, Osteoarthritis overview, *J. Geriatr. Nurs.* 22 (3) (2001) 124–131.
- [3] G.A. Hawker, L. Stewart, M.R. French, J. Cibere, J.M. Jordan, L. March, M. Suarez-Almazor, R. Gooberman-Hill, Understanding the pain experience in hip and knee osteoarthritis—an OARSI/OMERACT initiative, *Osteoarthr. Cartil.* 16 (4) (2008) 415–422.
- [4] E. Losina, A.M. Weinstein, W.M. Reichmann, S.A. Burbine, D.H. Solomon, M. Daigle, B.N. Rome, et al., Lifetime risk and age at diagnosis of symptomatic knee osteoarthritis in the US, *Arthritis Care Res.* 65 (5) (2013) 703–711.
- [5] Migliore, Alberto, Gianfranco Gigliucci, Liudmila Alekseeva, Sachin Avasthi, Ravendhara R. Bannuru, Xavier Chevalier, Thierry Conroyer et al. "Treat-to-target strategy for knee osteoarthritis. International technical expert panel consensus and good clinical practice statements." *Therapeutic advances in musculoskeletal disease* 11 (2019): 1759720X19893800.
- [6] E. Christodoulou, S. Moustakidis, N. Papandriano, D. Tsapopoulos, E. Papageorgiou, Exploring deep learning capabilities in knee osteoarthritis case study for classification, in: In 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), IEEE, 2019, pp. 1–6.
- [7] C.J. Mun, T.J. Speed, P.H. Finan, T.H. Wideman, P.J. Quartana, M.T. Smith, A Preliminary Examination of the Effects and Mechanisms of Cognitive Behavioral Therapy for Insomnia on Systemic Inflammation Among Patients with Knee Osteoarthritis, *Int. J. Behav. Med.* (2023) 1–10.
- [8] N. Kour, S. Gupta, S. Arora, A survey of knee osteoarthritis assessment based on gait, *Arch. Comput. Meth. Eng.* 28 (2021) 345–385.

- [9] J.H. Kellgren, JS1006995 Lawrence, Radiological assessment of osteo-arthrosis, *Ann. Rheum. Dis.* 16 (4) (1957) 494.
- [10] C. Kokkotis, S. Moustakidis, E. Papageorgiou, G. Giakas, D.E. Tsaopoulos, Machine learning in knee osteoarthritis: A review, *Osteoarthritis and Cartilage Open* 2 (3) (2020) 100069.
- [11] P.S. Yeoh, K.W. Qing, S.L. Lai, K.H. Goh, Y.C. Hum, Y.K. Tee, S. Dhanalakshmi, Emergence of deep learning in knee osteoarthritis diagnosis, *Comput. Intell. Neurosci.* (2021, 2021,) 1–20.
- [12] S. Aladhadh, R. Mahum, Knee Osteoarthritis Detection Using an Improved CenterNet With Pixel-Wise Voting Scheme, *IEEE Access* 11 (2023) 22283–22296, <https://doi.org/10.1109/ACCESS.2023.3247502>.
- [13] C. Kokkotis, C. Ntakolia, S. Moustakidis, G. Giakas, D. Tsaopoulos, Explainable ML for knee osteoarthritis diagnosis based on a novel fuzzy feature selection methodology, *Physical and Engineering Sciences in Medicine* 45 (1) (2022) 219–229.
- [14] C. Ma, Q. Wang, T. Ding, J. Yang, Knee Joint Pathology Screening Using Time-Domain Multidimensional Fusion Feature and Random Forest, in: In 2022 China Automation Congress (CAC), IEEE, 2022, pp. 2194–2199.
- [15] A. Rehman, A. Raza, F.S. Alamri, B. Alghofaily, T. Saba, Transfer Learning-Based Smart Features Engineering for Osteoarthritis Diagnosis from Knee X-ray Images, *IEEE Access* (2023).
- [16] A. Raza, T.-L. Phan, H.-C. Li, N. Van Hieu, T.T. Nghia, C.T.S. Ching, A Comparative Study of Machine Learning Classifiers for Enhancing Knee Osteoarthritis Diagnosis, *Information* 15 (4) (2024) 183.
- [17] K. Messaoudene, K. Harrar, Computerized diagnosis of knee osteoarthritis from x-ray images using combined texture features: Data from the osteoarthritis initiative, *Int. J. Imaging Syst. Technol.* 34 (2) (2024) e23063.
- [18] L.K. MunishKhanna, Singh, Hitendra Garg, A novel approach for human diseases prediction using nature inspired computing & machine learning approach, *Multimed. Tools Appl.* 83 (6) (2024) 17773–17809.
- [19] M. Khanna, N. Chauhan, D.K. Sharma, L.K. Singh, A Multi-Objective Approach for Test Suite Reduction During Testing of Web Applications: A Search-Based Approach, *International Journal of Applied Metaheuristic Computing (IJAMC)* 12 (3) (2021) 81–122.
- [20] H.T. Vedpal, N. Chauhan, M. Khanna, Test case prioritization using a Hybrid Chaotic Flower-fruit fly optimization algorithm with multiple objectives, *Multimed. Tools Appl.* 83 (10) (2024) 28395–28418.
- [21] L.K. Singh, K. Shrivastava, An enhanced efficient approach for feature selection for chronic human disease prediction: A breast cancer study, *Heliyon* (2024).
- [22] L.K. Singh, M. Khanna, H. Garg, R. Singh, Efficient feature selection based novel clinical decision support system for glaucoma prediction from retinal fundus images, *Med. Eng. Phys.* 123 (2024) 104077.
- [23] L.K. Singh, M. Khanna, H. Monga, G. Pandey, Nature-inspired algorithms-based optimal features selection strategy for COVID-19 detection using medical images, *N. Gener. Comput.* (2024) 1–64.
- [24] A. Sharma, P.K. Mishra, Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis, *Int. J. Inf. Technol.* (2022) 1–12.
- [25] N. Bayramoglu, M.T. Nieminen, S. Saarakkala, Machine learning based texture analysis of patella from X-rays for detecting patellofemoral osteoarthritis, *Int. J. Med. Inf.* 157 (2022) 104627.
- [26] U. Yunus, J. Amin, M. Sharif, M. Yasmin, S. Kadry, S. Krishnamoorthy, Recognition of knee osteoarthritis (KOA) using YOLOv2 and classification based on convolutional neural network, *Life* 12 (8) (2022) 1126.
- [27] A.M. Raisuddin, H.H. Nguyen, A. Tiulpin, Deep Semi-Supervised Active Learning for Knee Osteoarthritis Severity Grading, in: In 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), IEEE, 2022, pp. 1–5.
- [28] X. Wang, S. Liu, C. Zhou, Classification of Knee Osteoarthritis Based on Transfer Learning Model and Magnetic Resonance Images, in: In 2022 International Conference on Machine Learning, Control, and Robotics (MLCR), IEEE, 2022, pp. 67–71.
- [29] A.S. Mohammed, A.A. Hasanaath, G. Latif, A. Bashar, Knee osteoarthritis detection and severity classification using residual neural networks on preprocessed x-ray images, *Diagnostics* 13 (8) (2023) 1380.
- [30] A. El-Ghany, M.E. Sameh, A.A. Abd El-Aziz, A fully automatic fine tuned deep learning model for knee osteoarthritis detection and progression analysis, *Egyptian Informatics Journal* 24 (2) (2023) 229–240.
- [31] P. Chen, Knee osteoarthritis severity grading dataset, *Mendeley Data* 1 (2018) 21–23.
- [32] J.H. Kellgren, J.S. Lawrence, Radiological assessment of osteo-arthrosis, *Ann. Rheum. Dis.* 16 (4) (1957) 494–502.
- [33] M. Jigin, M.S. Madhulika, G.D. Divya, R.K. Meghana, S. Apoorva, Feature extraction using convolution neural networks (CNN) and deep learning, in: In 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE, 2018, pp. 2319–2323.
- [34] Feng, Han, Jianfei Li, and Ding-Xuan Zhou. "Approximation analysis of CNNs from feature extraction view." *arXiv preprint arXiv:2210.09041* (2022).
- [35] Karen, Simonyan. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv: 1409.1556* (2014).
- [36] H.M. Alshamlan, G.A.H. Badr, Y.A. Alohal, Genetic Bee Colony (GBC) algorithm: A new gene selection method for microarray cancer classification, *Comput. Biol. Chem.* 56 (2015) 49–60.
- [37] D. Karaboga, An Idea Based on Honey Bee Swarm for Numerical Optimization, *Computer Engineering Department, Erciyes University, Erciyes*, 2005.
- [38] Jabeen H, Jalil Z, Baig AR. Opposition based initialization in particle swarm optimization. In: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers; New York, NY, USA. 2009, pp. 2047–2052.
- [39] Reynolds, Craig W. "Flocks, herds and schools: A distributed behavioral model." In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pp. 25–34. 1987.
- [40] Heppner, Frank. "A stochastic nonlinear model for coordinated bird flocks." *The ubiquity of chaos* (1990).
- [41] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (1995) 273–297.
- [42] V. Jakkula, Tutorial on Support Vector Machine (svm) 37, no. 2(5) (2006) 3.
- [43] Tsocharidis, Ioannis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. "Support vector machine learning for interdependent and structured output spaces." In *Proceedings of the twenty-first international conference on Machine learning*, p. 104. 2004.
- [44] M.-L. Zhang, Z.-H. Zhou, ML-KNN: A lazy learning approach to multi-label learning, *Pattern Recogn.* 40 (7) (2007) 2038–2048.
- [45] Howe, Nicholas, and Claire Cardie. "Examining locally varying weights for nearest neighbor algorithms." In *Case-Based Reasoning Research and Development: Second International Conference on Case-Based Reasoning, ICCBR-97 Providence, RI, USA, July 25–27, 1997 Proceedings* 2, pp. 455–466. Springer Berlin Heidelberg, 1997.
- [46] Chomboon, Kittipong, Pasapitch Chujai, Pongsakorn Teerarassamee, Kittisak Kerdrasop, and Nittaya Kerdrasop. "An empirical study of distance metrics for k-nearest neighbor algorithm." In *Proceedings of the 3rd international conference on industrial application engineering*, vol. 2. 2015.
- [47] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [48] Hegde, Chiranth, Scott Wallace, and Ken Gray. "Using trees, bagging, and random forests to predict rate of penetration during drilling." In *SPE Middle East Intelligent Oil and Gas Symposium*, p. D011S001R003. SPE, 2015.
- [49] P. Xanthopoulos, P.M. Pardalos, T.B. Trafalis, P. Xanthopoulos, P.M. Pardalos, T. B. Trafalis, Linear discriminant analysis, *Robust Data Mining* (2013) 27–33.
- [50] J. Ye, R. Janardan, Q.i. Li, Two-dimensional linear discriminant analysis, *Adv. Neural Inf. Procs. Syst.* 17 (2004).
- [51] Bose, Anandh Sam Chandra, and C. Srinivasan. "Machine Learning based Detection and Severity Classification of Knee Osteoarthritis." In *2023 International Conference on Emerging Research in Computational Science (ICERCS)*, pp. 1–6. IEEE, 2023.
- [52] A. Qadir, R. Mahum, S. Aladhadh, A Robust Approach for Detection and Classification of KOA Based on BILSTM Network, *Comput. Syst. Sci. Eng.* 47, no. 2 (2023).
- [53] J. Yang, Q. Ji, M. Ni, G. Zhang, Y. Wang, Automatic assessment of knee osteoarthritis severity in portable devices based on deep learning, *J. Orthop. Surg. Res.* 17 (1) (2022) 540.
- [54] M.W. Brejnæbøl, P. Hansen, J.U. Nybing, R. Bachmann, U. Ratjen, I.V. Hansen, A. Lenskjold, M. Axelsen, M. Lundemann, M. Boesen, External validation of an artificial intelligence tool for radiographic knee osteoarthritis severity classification, *Eur. J. Radiol.* 150 (2022) 110249.
- [55] B. Muhammad, Mohammed, Mohammed Yesin, Interpretable and parameter optimized ensemble model for knee osteoarthritis assessment using radiographs, *Sci. Rep.* 11 (1) (2021) 14348.
- [56] A.D. Goswami, Automatic classification of the severity of knee osteoarthritis using enhanced image sharpening and cnn, *Appl. Sci.* 13 (3) (2023) 1658.