

Received January 21, 2020, accepted February 17, 2020, date of publication February 19, 2020, date of current version March 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975135

An Accurate Ensemble Classifier for Medical Volume Analysis: Phantom and Clinical PET Study

MHD SAEED SHARIF¹, MAYSAM ABBOD², ALI AL-BAYATTI³,
ABBES AMIRA³, (Senior Member, IEEE), AHMED S. ALFAKEEH⁴,
AND BAL SANGHERA⁵

¹School of Architecture, Computing and Engineering, ACE, UEL, London E16 2RD, U.K.

²Department of EC, College of Engineering, Design and Physical Sciences, Brunel University London, Uxbridge UB8 3PH, U.K.

³School of Computer Science and Informatics, CEM, De Montfort University, Leicester LE1 9BH, U.K.

⁴Department of Information Systems, Faculty of Computing and Information Systems, King Abdul Aziz University, Jeddah 21589, Saudi Arabia

⁵Paul Strickland Scanner Centre, Mount Vernon Hospital, London HA6 2RN, U.K.

Corresponding author: Mhd Saeed Sharif (s.sharif@uel.ac.uk)

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, under Grant DF 319611441.

ABSTRACT The predominant application of positron emission tomography (PET) in the field of oncology and radiotherapy and the significance of medical imaging research have led to an urgent need for effective approaches to PET volume analysis and the development of accurate and robust volume analysis techniques to support oncologists in their clinical practice, including diagnosis, arrangement of appropriate radiotherapy treatment, and evaluation of patients' response to therapy. This paper proposes an efficient optimized ensemble classifier to tackle the problem of analysis of squamous cell carcinoma in patient PET volumes. This optimized classifier is based on an artificial neural network (ANN), fuzzy C-means (FCM), an adaptive neuro-fuzzy inference system (ANFIS), K-means, and a self-organizing map (SOM). Four ensemble classifier machines are proposed in this study. The first three are built using a voting approach, an averaging technique, and weighted averaging, respectively. The fourth, novel ensemble classifier machine is based on the combination of a modified particle swarm optimization (PSO) approach and weighted averaging. Experimental National Electrical Manufacturers Association and International Electrotechnical Commission (NEMA IEC) body phantom and clinical PET studies of participants with laryngeal squamous cell carcinoma are used for the evaluation of the proposed approach. Superior results were achieved using the new optimized ensemble classifier when compared with the results from the investigated classifiers and the non-optimized ensemble classifiers. The proposed approach identified the region of interest class (tumor) with an average accuracy of 98.11% in clinical datasets of patients with laryngeal tumors. This system supports the expertise of clinicians in PET tumor analysis.

INDEX TERMS Medical imaging, tumor, committee machine, particle swarm optimization, squamous cell carcinoma.

I. INTRODUCTION

The investigation and analysis of the volume of positron emission tomography (PET) is crucial for different clinical and diagnosis procedures, such as decreasing noise, artifact evacuation, tumor evaluation in the management stage, and to help plan the appropriate radiotherapy treatment for patients [1]. PET is dynamically consolidated for the administration of patients. The outcomes of clinical investigations

using fluorodeoxyglucose (FDG)-PET have shown its advantage in the analysis, organization and assessment of patient reactions to treatment [2]–[4]. The use of cutting-edge elite programming evaluation methodologies is valuable in helping clinicians in clinical diagnosis and in arranging radiotherapy. In spite of the fact that therapeutic volume examination seems basic, in-depth knowledge of the organs and physiology is necessary to achieve such analysis from clinical restorative images. Basically, the clinical expert monitors each slice, delineates the borders from among the images, and thus characterizes every area. This is generally carried

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

out image by image (in 2D) for a 3D volume and requires a “re-slicing” of the area into the transaxial, sagittal, and coronal planes. Also, recognition of finer image features and the different changes are frequently required. Despite the fact that, for a commonplace 3D informational index, a manual, comprehensive clinical expert investigation can take a few hours to finish, this methodology is perhaps the most dependable and precise technique for restorative image examination. This is because of the monstrously multifaceted nature of the human visual framework, which is an appropriate framework for this task [5]–[8].

The systems based on a combination of classifier frameworks is usually effective in identifying different patterns and performing classification requirements. This can be achieved through various means, including combined decisions, different combinations of experts or specialists, ensemble classifiers, a fusion of different approaches, accord conglomerations, dynamic classifier determinations, composite classifier frameworks, half-and-half strategies, smart agents, framework supposition pools and board-of-committee ensemble machines [9]–[17]. The inspiration for such frameworks might come from a perception that particular classifiers are prevalent in various circumstances, or it might follow from the idea of the application in question. In addition, endeavors may be concentrated on improving the capacity for speculation and upgrading the precision of the classification.

Ensemble classifiers have different structures and are employed in a variety of applications [18]. Ensemble classifiers significantly outperformed other approaches used for data of microwave breast screening acquired in the clinical trial presented in [9]. Another study presented a neural network ensemble design for image classification purposes, but this work only used neural networks; the ensemble we developed, in contrast, deals with the image as whole and uses a combination of techniques [19]. Fusion of contextual information for the purpose of image recognition has been presented in [20], which shows that fusing the information can lead to better outcomes. Multicategory classification problems have been addressed using ensembles of binary classifiers in [21]. A gas-recognizing committee machine that joins different gas identification approaches to get a bound-together choice with improved precision was presented in [22]. The ensemble classifier was executed by amassing the yields of five gas-recognizing proof approaches through a cutting-edge casting ensemble with very good outcomes. A casting ensemble for spoken-influence classification was employed in [23], and the achieved committee precision was contrasted with the correctness of each separate classifier. In another study, a weighted casting committee machine was employed to recognize the human face and voice [24]. A hierarchical ensemble classifier was proposed in [25] based on multiple Fisher’s linear discriminant classifiers, where each one embodied different facial evidence for face recognition.

An ensemble classifier was used in [26] to identify tissue in black and white images. A group of multilayer perceptrons (forming a basic neural system) was set to learn data inputs comprising different surface patterns, and the data outputs comprised tissue type classes that were interpreted by clinical specialists. At that point, an ensemble classifier was developed via preparing a Bayesian classifier to combine the classification approaches of the neural systems. Results were contrasted among comparable AI-based techniques, such as support vector machine and multiclass Bayesian ensemble classifier. The designed methodology was used to identify pressure ulcers, a clinical pathology of localized harm to the skin and underlying tissue brought about by weight, shear force, or contact. A mean move methodology and a locale-developing procedure were used for productive region division. Analysis and treatment of pressure ulcers are exorbitantly expensive for health administrations. Correct assessment of injury is a basic task for ensuring the adequacy of treatment and care. Physicians normally assess each pressure ulcer by visual investigation of the affected tissues, which is an inadequate methodology for assessing the extent of the injury.

An ensemble classifier of neural systems was presented in a paper identifying the masses found in mammograms [27]. This ensemble was employed to group masses into two classes, malignant and benign. This study used 20 areas of intrigue identified with harmful tumors and 37 others belonging to benign tumors. A set of multilayer perceptrons was used as a complete ensemble of neural systems. The outcomes were achieved by consolidating the reactions of the individual classifiers. The research study proposed in [28] investigated several AI techniques for the purpose of identifying harmful and benign bunched microcalcifications. This study’s kernel-based approach methodology accomplished a performance of 85%.

A committee machine of neural systems intended to enhance the precision and vigor of identifying samples of gene information was developed in [29]. Another committee machine, based on a voting framework for identifying multi-class protein creases, was presented in [30]. These studies are crucial because identifying the protein structure is critical for knowing the relationship between sequence (i.e., structures) and conceivable protein functions.

The study presented in this paper investigates the efficiency of different committee machines and proposes a novel optimized committee machine to tackle the problem of analysis of squamous cell carcinoma in patient PET volumes. Our optimized classifier is based on an artificial neural network (ANN), fuzzy C-means (FCM), an adaptive neuro-fuzzy inference system (ANFIS), K-means, and a self-organizing map (SOM). This study includes four committee machines. The first is based on a voting approach, where every individual classifier generates a specific outcome. The second is based on an averaging technique, where the class outcome revealing the biggest average weight is

selected as the most accurate. The third is based on weighted averaging, where the generated outcomes from all included techniques are timed with the archived predicted weights. The fourth and novel committee machine depends on the combination of a modified particle swarm optimization (PSO) approach and weighted averaging. The proposed optimized committee machine is evaluated using experimental National Electrical Manufacturers Association and International Electrotechnical Commission (NEMA IEC) body phantom and clinical PET studies of seven participants diagnosed with laryngeal squamous cell carcinoma. Very promising outcomes are achieved using the new optimized committee machine (CM4), as illustrated in the following sections.

II. THEORETICAL BACKGROUND

A. VOTING-BASED ENSEMBLE

The voting technique is one of the more popular methodologies for consolidating the outcomes of different classifiers. In this method, every single classifier generates a decision rather than a weight. The chosen class is the one most commonly selected by the various classifiers. Hence, the yield forecast (V_p) is resolved as follows:

$$V_p = \begin{cases} x_1 & \text{when } \sum_{i=1}^K x_i > T \\ x_2 & \text{when } \sum_{i=1}^K x_i < T \\ \text{tie} & \text{when } \sum_{i=1}^K x_i = T \end{cases} \quad (1)$$

where K is the quantity of classifiers and T is a turning limit. If 50% of the involved classifiers vote in favor of one class and the remaining 50% vote in favor of the other class, then a tie status is reached. This occurs when an evenly divisible number of classifiers is employed in the ensemble. Nevertheless, in the proposed ensemble, an odd number of classifiers is conveyed to stay away from this issue. Moreover, the most well-known technique among median, least and greatest techniques is the preponderance vote technique [31]–[33].

B. AVERAGING-BASED ENSEMBLE

The averaging-based ensemble carries out an averaging technique on the outcome of each classifier for every representative class across the whole ensemble. The class outcome with the largest amount is then selected. The outcome is shown in (2):

$$Q(x) = \operatorname{argmax}_{j=1\dots N} \left(\frac{1}{K} \sum_{i=1}^K y_{ij}(x) \right) \quad (2)$$

where N is the classes quantity, $y_{ij}(x)$ is the outcome value of the i^{th} classifier for the j^{th} class of the input x , and K is the number of classifiers used in the whole ensemble [32].

C. WEIGHTED AVERAGING ENSEMBLE

The weighted averaging technique is similar to the averaging technique described above, with the additional parameter that the classifiers' outcomes are multiplied by the

archived predicted weights. The outcome is shown in following equation [32]:

$$Q(x) = \operatorname{argmax}_{j=1\dots N} \left(\frac{\sum_{i=1}^K w_i y_{ij}(x)}{\sum_{i=1}^K w_i} \right) \quad (3)$$

The weights w_i where $i = 1, \dots, K$, are derived by minimizing the errors of the different classifiers in the training group. In this study, which uses PET as the application, the prediction accuracy of each classifier for each corresponding class has been used as a weight for that class. The desired output of each classifier y_{di} can be written as the actual output y_i plus an error e_i , as shown in (4):

$$y_{di}(x) = y_i(x) + e_i(x) \quad (4)$$

D. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) is set up based on social swarm behavior, which can control convergence [34]–[36]. Social swarm behavior can be portrayed as follows: Consider, for example, a flock of birds with the goal of locating a warm place to travel to. Having no earlier knowledge of such a place, the birds set out in arbitrary directions with irregular speeds, searching for a desirable spot. Each bird can remember its own discovered area and one way, or another also knows the headings where different members of the flock also found a legitimate spot. A hesitant bird, caught between the direction it found and those found by others, accelerates in both directions, thus modifying its bearing to fly somewhere between the two known headings. During flight, this bird may locate an even better place than the one it found before. It would then be attracted to this new area as well as to the other "best" area found by the entire flock. Occasionally, one bird may fly in a better direction than had been encountered by any other member of the flock. The entire flock would then be drawn toward that specific area in addition to their own find. Eventually, flock's collective flight guides them to the best location they are searching for [37], [38].

PSO has many advantages, such as its algorithmic simplicity. Moreover, it has one straightforward operator, which is velocity. This feature decreases computational time and complexity. In PSO, a defined group of elements must be chosen and carefully controlled according to the application in question. These parameters (elements) are experimentally investigated and optimized based on the proposed application.

III. METHODS AND MATERIALS

A. PHANTOM AND CLINICAL STUDIES

1) PHANTOM STUDIES DATA

The main informational index used in this research investigation was gathered from the NEMA IEC image quality body phantom. This phantom comprises an elliptical water-filled cavity which contains six spherical inserts, suspended by plastic rods, of volumes 0.5, 1.2, 2.6, 5.6, 11.5, and 26.5 ml. The internal diameters of these spheres are 10, 13, 17, 22, 28, and 37 mm, respectively. The volume of the PET has a size of

$168 \times 168 \times 66$ voxels and each voxel has measurements of $4.07 \times 4.07 \times 5 \text{ mm}^3$, equal to a voxel volume of 0.0828 ml . This phantom has been extensively used in the literature for evaluation of image quality and for validation of quantitative processes [39]–[42]. PET emission data was reconstructed using a CT-based attenuation correction performed after Fourier rebinning and model-based scatter correction. The PET volumes were reproduced using two-dimensional iterative standardized normalized attenuation weighted ordered subsets expectation-maximization (NAW-OSEM). In this experiment, the following default parameters were employed: ordered OSEM repetitive reconstruction, four recurrences with eight subsets, followed by a post-processing Gaussian filter (5 mm) [43]. Phantom volumes were obtained using a GE DST clinical PET-CT scanner.

2) CLINICAL STUDIES DATA

Our clinical study focuses on a specific type of cancer data, hence, the clinical dataset in this research consisted of PET images from seven patients with T3–T4 laryngeal squamous cell carcinoma. *T3* represents a tumor in the larynx that has rendered one of the vocal cords incapable of movement, and *T4* represents a tumor that has extended beyond the larynx. Prior to treatment, every patient underwent an FDG-PET examination. Patients were immobilized with a customized thermoplastic mask attached to a flat tabletop to prevent complex neck motions. The procedure was as follows: a 10-min transmission scan was performed using the Siemens Exact HR camera (CTI, Knoxville, USA). Afterwards, a 1-h dynamic 3D emission scan was performed immediately after intravenous infusion of 185–370 MBq (5–10 mCi) of FDG. This scan has eight frames with variable span running from 90 to 600 s. All images were corrected for dead time, random, scatter, attenuation and decay and then reconstructed using a 3D OSEM algorithm. The ground truth evaluation of the tumor was based on the knowledge of expert clinicians who used their training and experience to identify suspected sites of disease through quantification measurement and visual assessment, while the histology (through biopsy) confirmed whether a disease was present at individual suspected sites and characterized its physical distribution and extent in biopsied tissue accordingly. The volume of this dataset was $128 \times 128 \times 47$ voxels for each participant, with a voxel size of $2.17 \times 2.17 \times 3.13 \text{ mm}^3$. The result was a total of 329 images to process [44]–[46].

B. THE OPTIMIZATION APPROACH

We developed an optimization algorithm for this study based on the PSO approach [34]. It can be summarized as follows:

1. The initial step of implementing PSO is to choose the parameters and characterize the search range for each of them.

2. The mean squared error fitness function is chosen to show the goodness of fit of the optimization solution. The PSO particles are optimized by the fitness function, which is formulated as the objective function.

3. Each molecule starts at its own random location with an arbitrary speed searching for the ideal position in the arrangement space. As the initial position of each molecule is the only location encountered at the beginning, this position becomes respective of each molecule (*pbest*). Each molecule has its own *pbest* dictated by the path that it has taken.

4. The first global best solution found by the rest of the swarm (*gbest*) is then selected from among these initial positions. From that point onward, the methodology moves every molecule separately by a small amount through the whole swarm and thinks about *gbest* and *pbest*.

5. The particle's speed is imperative in the optimization system. The speed of the molecule is changed by the relative locations of *pbest* and *gbest*. Every molecule is processed as a point in the D dimensional defined space. The i^{th} molecule can be shown as $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$.

The best previous position (the position which gave the best fitness value) of the i^{th} molecule is monitored and can be shown as $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})$. An index of the best molecule among the whole populace is represented by the symbol g . The range of velocity variation for particle i can be shown as $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$. This velocity is updated according to the following equation:

$$\begin{aligned} v_{ip}(q+1) &= \omega v_{ip}(q) + c_1 \cdot \text{rand}(1) \\ &\quad \cdot (p_{ip}(q) - x_{ip}(q)) + c_2 \cdot \text{rand}(2) \cdot (p_{gp}(q) - x_{ip}(q)) \end{aligned} \quad (5)$$

where $0 \leq i \leq (n-1)$, $1 \leq p \leq D$, n is the number of particles in a group and D is the dimension of the search space. For each particle, there are D number of parameters that are used to identify the particle location in the search space. For a specific particle, q is the repetition indicator, $v_{ip}(q)$ is the speed of particle i at repetition q and $\text{rand}(1)$ and $\text{rand}(2)$ are random values within $[0, 1]$. The inactivity weight factor is ω , which determines to what degree the molecule stays along its unique course unaltered by the draw of *gbest* and *pbest*. The quickening constants are c_1 , and c_2 , where c_1 is a factor determining how much the molecule is impacted by the memory of its best location and c_2 is a factor determining how much the molecule is affected by the remainder of the particles in the group.

6. Based on the previous step and after updating the velocity, the new location, to which the particle moved, is updated based on the following:

$$x_{ip}(q+1) = x_{ip}(q) + v_{ip}(q+1) \quad (6)$$

where x_{ip} is the available location of a certain particle i at a repetition q .

7. Afterwards, the process from step 4 onward is repeated. Repetition continues until the end model condition is achieved.

C. THE ENSEMBLE OPTIMIZATION APPROACH

Based on the understanding of the conceptual basis of the PSO, we developed an algorithmic methodology as a

novel enhancement approach method for PET applications, to improve the performance of the developed ensemble. The PET optimization approach can be summarized as follows:

1. Every particle in the swarm accelerates toward the best overall position and best settings while constantly checking the value of its current location.

To implement this approach, a specific procedure is followed: First, we need to select then optimize five different parameters—one value for every proposed classifier—and give them a reasonable range in which to search for the optimal solution. Based on the initial experiments completed for the application of PET volumes, the best-chosen values for all the processed datasets is [0, 1.5]. The optimized values (R_1, R_2, R_3, R_4, R_5) were used within the developed ensemble in order to obtain the optimal outcome based the following equation:

$$Q(x) = \underset{j=1 \dots N}{\operatorname{argmax}} \left(\frac{\sum_{i=1}^K R_i \operatorname{Acc}_i y_{ij}(x)}{\sum_{i=1}^K \operatorname{Acc}_i} \right) \quad (7)$$

where Acc_i is the classification accuracy of a particular classifier i and N is the number of classes processed.

2. The second essential step is to evaluate the quality of the achieved solution. The accuracy function was chosen to provide a single figure referring to the effectiveness of the provided solution. The performance index of this function can be calculated as follows:

$$PI = 1 - \frac{1}{N} \sum_{i=1}^K \frac{tp_i}{tp_i + fp_i + fn_i} \quad (8)$$

where PI is the performance index, fp is false positive, tp is true positive and fn is false negative. These values are extracted from the confusion matrix.

3. Every molecule has its own $pbest$ controlled by the way that it has experienced. The first global best solution found by the remainder of the swarm ($gbest$) is then chosen from among these initial positions.

4. Each particle is moved individually by a small amount through the entire swarm, and the $pbest$ and $gbest$ are compared. The accuracy function returns a value to be assigned to the current location. If that value is greater than the value at the respective $pbest$ for that particle, or the global $gbest$, then the appropriate locations are replaced with the current locations.

5. The particle velocity is calculated based on (5). We have introduced the random parameter in Equation 5 to imitate the lightly unforeseeable behavior element of swarm in nature. Variant empirical parametric studies have been performed in the literature to determine the optimal amount of the c_1 with c_2 . It has been determined before that the best choice for both c_1 combined with c_2 is 2.0 [47], [48]. But, the first round of experiments performed using every single data sets of PET shows that the optimal amount of w is 0.7298 and the best amount of c_1 , and c_2 is 1.49618. The obtained experimental values are achieved throughout the following experimental approach, where compels conditions have been connected to these elements, for example, constraining the w

within the values of [0. . . 1]. The proposed approach has set the index performance to a value of 0.01. The proposed algorithm has saved, compared, and choose the best parameters to be deployed:

$$c_1 = c_2 = 1.49618$$

$$w = 0.7298$$

1: Determine the constraints for each parameter

2: Initialize c_1

3: Initialize c_2

4: Initialize w

5: As long as $PI > 0.01$, do

6: $c_1 = c_1 + c_1 * \operatorname{rand}(\pm 0.1)$

7: $c_2 = c_2 + c_2 * \operatorname{rand}(\pm 0.1)$

8: Calculate $w = w + w * \operatorname{rand}(0.1)$

9: Import PSO Approach

10: Save the achieved parameters in an array

11: Finish the cycle

12: Evaluate the parameters and deploy the best ones.

where $\operatorname{rand}(0.1)$ was set in the range [0, 0.1].

6. The new location for each particle based on (6) is calculated.

7. The proposed approach performed a significant number of experiments, which determined a termination criterion of 100 iterations for each PET dataset. Where the maximum number of iterations is too large, the algorithm may become idle waiting for a change in the constant parameters; on the other hand, having too few iterations could result in the swarm having insufficient time to adequately explore the solution space and find the best solution.

8. A significant number of experiments were performed using both the phantom and clinical datasets to determine the best relationship between the particle number (N_p) and the maximum number of iterations (N_{it}). Hence, it was determined that the complete analysis of such correlation may be described by (9), applicable to both phantom and clinical datasets:

$$N_{it} \cong \frac{3}{2} N_p \quad (9)$$

Moreover, choosing the right number of particles is very important for improving the performance of the proposed approach. Therefore, these initial experiments have evaluated the most suitable number of particles for the datasets in question. The achieved results are discussed section IV.

D. DESCRIPTION OF THE OPTIMIZED PROPOSED SYSTEM

The proposed medical volume ensemble classifier for the analysis and classification of PET images is shown in Fig 1. We used practical NEMA IEC phantom and pharyngolaryngeal squamous cell carcinoma datasets to assess the performance of the developed optimized approach. The initial five classifiers were applied to the pre-processed PET images obtained from the scanner as follows:

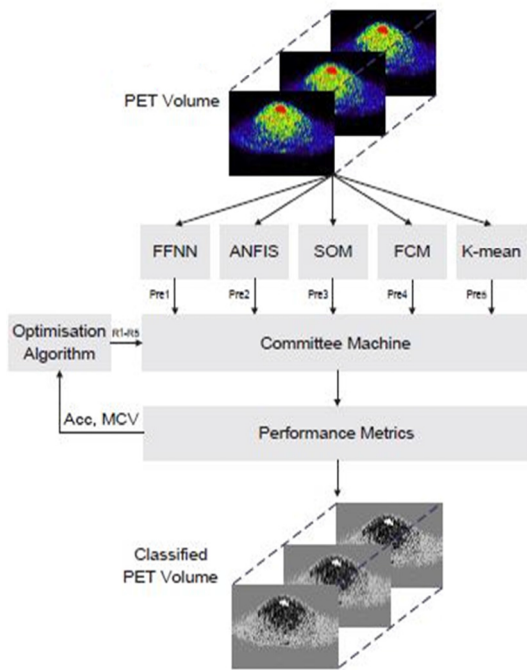


FIGURE 1. Novel Optimized Ensemble Classifier for PET Volume Classification.

1) FEEDFORWARD NEURAL NETWORK (FFNN)

The deployed FFNN has three layers (input, hidden and output). The node of its output is modeled as follows:

$$Y(j) = f \left[\sum_{i=1}^R w_{1,i}(j) \cdot p_i(j) + b \right] \quad (10)$$

In this model, tangent-sigmoid transfer function (f) is used for the input vector p_i , where b is the bias and w_1 and r are the weights. To achieve an output consistent with the training examples, these weights are updated through the following equation.

$$w_{1,i}(j+1) = w_{1,i}(j) + \alpha * p_i(j) * e(j) \quad (11)$$

where, based on the experiment, the learning rate (α) is determined to be 0.91, and $e(j)$ is the error calculated by taking the difference between the desired output Y_d and the actual output Y :

$$e(j) = Y_d(j) - Y(j) \quad (12)$$

Different parameters have been explored to achieve an appropriate architecture for an artificial neural network suitable for the PET application. These parameters include the training techniques; the Levenberg-Marquardt backpropagation training procedure was selected, determining the number of the hidden layers employed in the proposed network architecture and determining the number of hidden neurons (19) in every single layer [6]. One thousand (1,000) iterations were used during the training process. Repeated experiments were performed, with ten cycles associated with every architecture of the network. Afterwards, the optimal architecture was selected and trained.

2) ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (ANFIS)

The ANFIS model can learn input-output mapping based on human knowledge, which is provided in the form of “if-then” fuzzy rules. The fuzzy architecture is characterized by a set of rules, which are properly initialized and tuned by a learning algorithm. The rules are in the following form:

Rule 1: If (x is A_1) and (y is B_1) then ($f_1 = p_1x + q_1y + r_1$)

Rule 2: If (x is A_2) and (y is B_2) then ($f_2 = p_2x + q_2y + r_2$)

where x and y are the inputs, A_i and B_i are the fuzzy sets, f_i are the outputs within the fuzzy region specified by the fuzzy rule, p_i , q_i , and r_i are the design parameters that are determined during the training process.

Iterative tests were performed to assess the most appropriate parameters for ANFIS approach [10]. The experiment showed that following selected elements can achieve and generate both optimal performance and results. The influence value is set to 0.1, the accept ratio is set to 0.1, the squash element is determined at 0.25, and the reject ratio is set at 0.0015.

3) SELF-ORGANIZING MAP (SOM) [49]

For this PET application, the learning rate was chosen at a value of 0.6, and the number of training iterations was set to 1000.

4) FUZZY C-MEANS (FCM) [50]

The experiments determined the following parameters to achieve the FCM convergence, which suits the proposed PET application: the number of iterations is determined at a value of 500, the value m is equal to 2 and the least value for improvement is set at a value of $1e-5$.

5) K-MEANS [51]

This algorithm classifies n voxels into K clusters (where $K < n$). This algorithm chooses the number of classes, K , then randomly generates K classes and determines the cluster centers.

Collectively, these five classifiers provide the following classification predications (Pre): Pre1, Pre2, Pre3, Pre4 and Pre5, respectively.

The target class is a tumor (class 5) in the clinical datasets and a sphere (class 4) in the phantom datasets (simulated tumor). Each class refers to a different structure within the processed datasets. The selection of the appropriate number of classes number (N) related to each dataset in the analyzed PET images was made by experimenting and evaluating different values of N . The optimal value of N is determined based on the Bayesian information criterion (BIC) approach. BIC has gained notoriety as a significant approach for model selection and has been used in contexts varying from image processing and analysis to biological and sociological research. The BIC values were determined incrementally against increasing values of N . Values of N were chosen in the range from 2 to 8, since in this medical PET application, any additional separation is unnecessary, based

on analysis by medical experts. BIC values tend to increase indefinitely as the quantity of components increases in the model. An increase in BIC value indicates an improved model fit; however, these values typically stabilize on an approximate curve plateau, the beginning of which is usually taken to indicate the optimal N value for each dataset. Plotting BIC values against N for the phantom dataset showed that the optimal N is 4. However, the optimal N value for the clinical datasets is 5 classes for each patient, where class number 5 refers to the region of interest (tumor), while the remaining classes are the various other structures presented in the image [52]–[55].

In an initial stage, the generated outcomes from the deployed classifiers were fed to three ensemble classifiers (also called committee machines). The first ensemble (CM1) combines the five classifier outcomes by applying the voting approach. The second ensemble (CM2) combines the five outcomes using the averaging approach, while the third ensemble (CM3) combines the five outcomes using weighted averaging. After the initial stage using these three different ensemble classifiers, a novel approach for optimization built using the PSO approach was introduced and combined with the existing ensemble to improve the outcome and enhance the overall accuracy of the classification. The developed optimizer intensively searched for the most appropriate values for the optimization procedure (R1, R2, R3, R4, R5) of each classifier. These best values are then sent to and employed in the new ensemble, which in turns creates the optimized predication for the classification as well as the overall accuracy associated with each dataset. The optimized ensemble/committee machine (CM4) benefits from the mixture of supervised and unsupervised classifiers as well as from the proposed optimization approach. The most appropriate five parameters deployed in the new ensemble CM4 enhanced the overall performance as well as the accuracy achieved by the new CM4, as discussed in section IV. Each optimization parameter has the ability to represent the most appropriate solution related to its own classifier, so long as it was given suitable range of values. At the end of the optimization process, a performance pointer is created. This pointer is employed to determine the performance of the optimized ensemble CM4. The outputs are evaluated at the next stage, where a misclassification value (MCV), confusion matrix, and accuracy (Acc) are employed to assess the ensemble performance. The MCV represents the number of samples which are wrongly classified over the whole number of samples.

The novel ensemble classifier introduced here has generated significant results for all processed datasets, both phantom and clinical. The developed ensemble generated an accuracy as high as 99.9% for some clinical datasets. Section IV illustrates the results generated by the new ensemble system.

IV. RESULTS

The results are organized into two main sections for each type of dataset (phantom and clinical): The first section analyzes

TABLE 1. Prediction analysis using a 4×4 confusion matrix for CM1, CM2 and CM3 of phantom dataset outcomes and the accuracy (Acc) for each class.

CM	Class	C11	C12	C13	C14	Acc
CM1	C11	24939	-	-	-	0.9951
	C12	27	260	-	-	0.8333
	C13	85	25	2737	-	0.9473
	C14	10	-	42	99	0.6556
CM2	C11	27939	-	-	-	0.9990
	C12	24	263	-	-	0.7013
	C13	-	88	2759	-	0.9517
	C14	-	-	52	99	0.6556
CM3	C11	24939	-	-	-	0.9990
	C12	23	264	-	-	0.6168
	C13	-	141	2706	-	0.9337
	C14	-	-	51	100	0.6622

the results from the first three committee machines, CM1, CM2 and CM3. The second sub-section discusses the results from the optimized committee machine, CM4, with a focus on the accuracy of region of interest class (tumor).

A. PHANTOM DATASET

1) COMMITTEE MACHINE RESULTS

The MCV, confusion matrix, and Acc are employed to assess the performance of the developed ensembles. Table 1 shows the confusion matrix and Acc for the ensembles CM1, CM2 and CM3. The confusion matrix for the outputs of the first ensemble CM1 shows the following classification details: All class 1 voxels were accurately classified, 27 voxels related to class 2 were misclassified into the class 1, 85 voxels from class 3 were misclassified into class 1, and 25 voxels into class 2. Class 4, which represents the simulated tumor, had 10 voxels misclassified in class 1, and the other 42 voxels were misclassified in class 3. For the region of interest, 99 voxels were accurately classified. The confusion matrix for the second ensemble, CM2, shows that the following numbers of voxels were misclassified: 24 voxels from class 2 into class 1, 88 from class 3 into class 2, and 52 from class 4 into class 3, while class 1 was correctly classified.

Evaluating the outputs of the third committee machine CM3 based on the confusion matrix shows that the following numbers of voxels were misclassified: 23 voxels from class 2 into class 1, 141 from class 3 into class 2, and 51 from class 4 into class 3, while class 1 was correctly classified. Among the three ensembles, the best Acc was achieved by CM3.

2) OPTIMIZED CM RESULTS

Following the evaluation metrics of CM1, CM2 and CM3 presented in the previous section, the best performance is achieved by CM3. However, higher classification accuracy

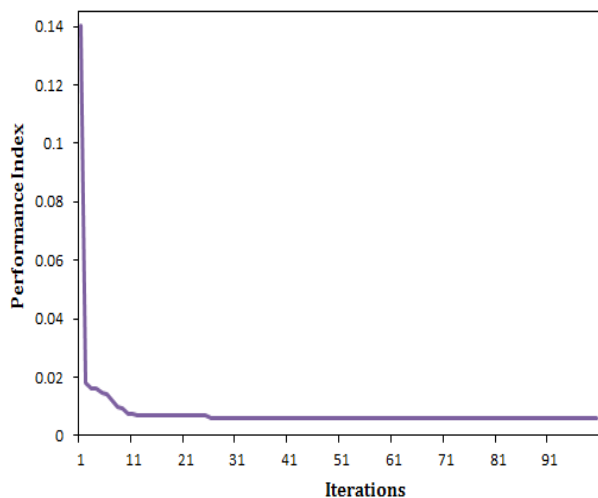


FIGURE 2. The optimization error model for the phantom dataset.

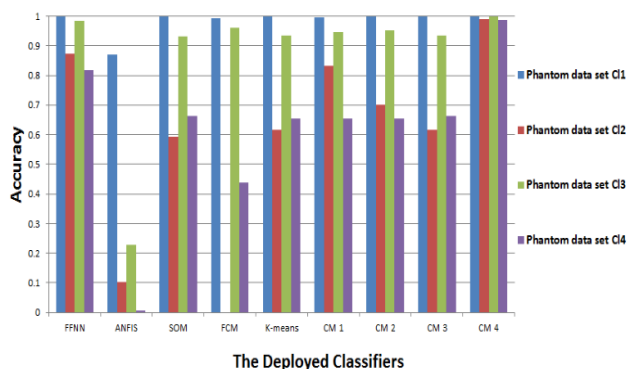


FIGURE 3. A comparison between the new proposed ensemble classifier, CM4, and the remaining approaches (FFNN, ANFIS, SOM, FCM, K-means, CM1, CM2 and CM3) based on the Acc for the phantom dataset.

and better performance are still required. Therefore, the optimized CM4 we developed was deployed within the system to process all the datasets. The initial experiments deploying the phantom dataset show that the most appropriate particles number required for the phantom dataset was 66 and was associated with 100 training iterations. This value is consistent with the one generated in (9). Once the training is performed, the optimization error/performance index achieved is $PI = 0.0061$. Fig. 2 shows the error obtained during the optimization process with 100 iterations, which had stabilized performance with no further reduction in error.

The experimental phantom dataset is then processed through the new optimization approach (CM4). All region of interest and class 2 voxels were accurately classified, 3 voxels from class 1 were misclassified into class 2 and only 2 voxels from class 3 were wrongly classified into class 4. The MCV of 0.0002 was the closest to “0” among all the other classifiers. The MCV values for all classifiers are presented in Table 4.

Significant improvement in accuracy was achieved using CM4 to detect and accurately classify the region of interest (simulated tumor). Fig. 3 shows a comparison between the new proposed ensemble classifier, CM4, and the remaining

TABLE 2. Prediction analysis using a 5 × 5 confusion matrix for CM1, CM2, and CM3 outcomes, and Acc score, for clinical dataset from Pt 1 (laryngeal tumor).

CM	Class	CI1	CI2	CI3	CI4	CI5	Acc
CM1	CI1	62	737	-	-	-	0.0172
	CI2	0	9353	-	-	-	0.9204
	CI3	1968	71	2790	303	-	0.5436
	CI4	660	-	-	162	-	0.1440
	CI5	173	-	-	-	105	0.3776
CM2	CI1	19	780	-	-	-	0.0130
	CI2	-	9353	-	-	-	0.9150
	CI3	-	88	3535	1509	-	0.6888
	CI4	495	-	-	327	-	0.1402
	CI5	157	-	-	-	121	0.4352
CM3	CI1	530	269	-	-	-	0.6633
	CI2	-	9353	-	-	-	0.9644
	CI3	-	76	4030	1026	-	0.7852
	CI4	-	-	-	563	259	0.2969
	CI5	-	-	-	48	230	0.4283

approaches (FFNN, ANFIS, SOM, FCM, K-means, CM1, CM2 and CM3) based on the Acc for the phantom dataset. The simulated tumor (sphere) represented by class 4 (C14) was accurately classified using CM4.

B. CLINICAL DATASET

1) COMMITTEE MACHINE RESULTS

The proposed ensemble classifier’s performance was evaluated through different assessment and analysis metrics for the clinical datasets from patients 1–7 (Pt 1–7). The confusion matrix illustrates the following results related to the ensemble CM1 outcomes of the data related to Pt 1 as an example of the clinical datasets for patients with pharyngolaryngeal squamous cell carcinoma: all class 2 voxels were accurately classified; 62 voxels from class 1 were accurately classified; 2790, 162, and 105 voxels related to classes 3, 4, and 5, respectively, are correctly classified. Table 2 lists the results obtained by the confusion matrix and Acc for each class of the dataset from Pt 1.

The confusion matrix illustrates the following results related to the ensemble CM2 outcomes of the data related to Pt 1: 780 voxels from class 1 were misclassified in class 2, 88 from class 3 were misclassified in class 2, 1509 from class 3 were misclassified in class 4, 495 from class 4 were misclassified in class 1, and 157 from class 5 were misclassified in class 1. Class 2 was accurately classified, as illustrated in Table 2.

The ensemble CM3 accurately classified both the voxels in class 2 as well as 230 voxels out of 278 voxels of the region of interest, while CM1 and CM2 classified only

TABLE 3. Prediction analysis using a 5 × 5 confusion matrix for CM1, CM2, and CM3 outcomes, and Acc score, for clinical dataset from Pt 2 (laryngeal tumor).

CM	Class	CI1	CI2	CI3	CI4	CI5	Acc
CM1	C11	53	2419	-	-	-	0.0123
	C12	0	8715	-	-	-	0.6661
	C13	53	1948	1213	-	-	0.3177
	C14	1022	-	-	5	-	0.0042
	C15	129	-	-	147	77	0.2181
CM2	C11	16	2456	-	-	-	0.0059
	C12	-	8715	-	-	-	0.6392
	C13	-	2463	1354	-	-	0.2861
	C14	105	-	914	8	-	0.0068
	C15	118	-	-	145	90	0.2549
CM3	C11	1931	541	-	-	-	0.7811
	C12	-	8715	-	-	-	0.7697
	C13	-	2066	1751	-	-	0.3735
	C14	-	-	871	156	-	0.1142
	C15	-	-	-	339	14	0.0396

105 and 121 voxels accurately, respectively. The ensemble CM3 also achieved an accuracy for class 1 that was superior to the ensembles CM1 and CM2. The detailed accuracy and confusion matrix are listed in Table 2.

Similar results were achieved for Pt 2; the details of the confusion matrix for CM (1–3) and the Acc of each class are presented in Table 3. In contrast, for Pt 3, the accuracies achieved by ensemble CM3 for classes 1 and 4 were 0.9225 and 0.8766, respectively. These values were better than the ones obtained from CM1 and CM2. In the dataset from Pt 4, class 2 was detected correctly by CM1, CM2 and CM3. For the dataset from Pt 5, CM3 detected all the classes; however, CM1 and CM2 detected only classes 2, 4 and 5. CM3 generated an accuracy of 1 (100%) for class 1 in the dataset from Pt 6; in contrast, the accuracy obtained by CM1 and CM2 are 0.0267 and 0.0116, respectively. In addition, the accuracy of CM3 for class 5 is the best among the first-stage committee machines (CM1, CM2 and CM3). For Pt 7, CM3 generated an accuracy of 1 (100%) for class 1. The voxels from class 2 are all correctly classified through the ensembles CM1, CM2 and CM3; and only ensemble CM1 was not able to identify class 3 voxels.

2) OPTIMIZED CM RESULTS

The patient datasets were also processed using the new optimized approach, CM4. Comprehensive experiments were performed to make sure the most appropriate optimization elements were employed to effectively classify the datasets in question. The optimum particle number required for all patient datasets was 66 particles with 100 training

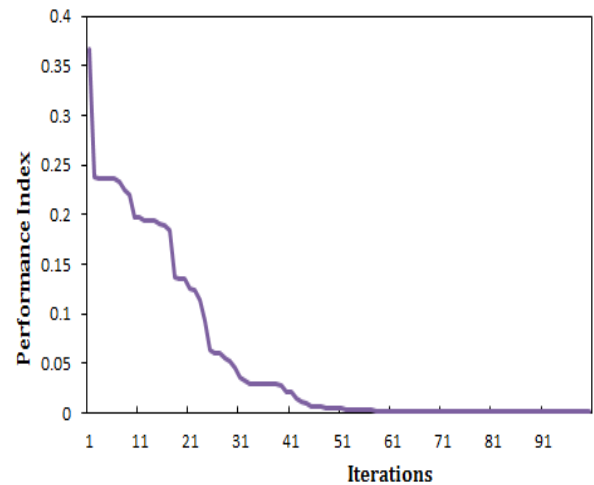


FIGURE 4. The optimization error model for clinical datasets.

iterations. Once the training procedure was complete, the error/performance indicator for the patient datasets was $PI = 0.0005$. These optimization parameters were used for the seven patient datasets of pharyngolaryngeal squamous cell carcinoma. Fig 4 shows the performance index achieved for the optimization procedure, which was associated with 100 iterations. The achieved model was generalized and validated to fit all the patient datasets. This model illustrated a stable robust performance to analyze all the datasets with a stabilized index performance. As stated at the beginning of the section, the little number of the misclassification case is due to the difficult nature of the processed datasets and due to the fact that any developed classification approach can have a certain level of misclassification, which varies based on the type of the approach and its parameters.

In the dataset for Pt 1, the processed voxels in classes 1, 2, 4 and 5 were accurately identified, and only 6 voxels related to class 3 were misclassified in class 4. In the dataset for Pt 2, in contrast, classes 1 and 4 were correctly classified; however, there were 2 voxels related to class 2 that were misclassified in classes 1 and 4, three voxels related to class 3 misclassified in class 1, and only 1 voxel related to class 5 misclassified into class 1. In the dataset from Pt 3, all voxels in class 2 were accurately allocated, while only 3 voxels related to the region of interest were misclassified to class 1. For Pt 4, the results were similar, as 5 voxels out of the 1046 voxels in the region of interest were misclassified in class 4. In the dataset from Pt 5, only 1 voxel out of 1174 in the region of interest was misclassified to class 1. For Pt 6, in the region of interest, classes 1 and 3 voxels were accurately classified, while 3 voxels related to class 2 were misclassified into class 1 and 2 voxels related to class 4 were misclassified into class 3.

The accuracy achieved for all the classes (1–5) using the developed CM4 is very satisfactory, as shown in Fig. 5. CM4 accurately classified not only the region of interest (tumor) but also all the classes in the clinical datasets (Pt 1–7). This fact shows the robustness of the developed

TABLE 4. MCV values for all the approaches and ensembles (CM1, CM2, CM3 and CM4) of the phantom dataset and clinical datasets (Pts 1–7).

Dataset	Phantom	Pt 1	Pt 2	Pt 3	Pt 4	Pt 5	Pt 6	Pt 7
Classifier	MCV							
FFNN	0.0024	0.5692	0.5000	0.4169	0.4662	0.4037	0.5050	0.4741
ANFIS	0.1384	0.3302	0.3073	0.5333	0.4421	0.5104	0.5539	0.5316
SOM	0.0080	0.1887	0.3807	0.7536	0.4680	0.2529	0.6279	0.4867
FCM	0.0139	0.3042	0.2692	0.7891	0.3474	0.2811	0.4064	0.4542
K-means	0.0076	0.1881	0.3808	0.7548	0.4534	0.2535	0.6271	0.4867
CM1	0.0066	0.2387	0.3858	0.7621	0.4331	0.2550	0.5993	0.4702
CM2	0.0058	0.1848	0.3784	0.7174	0.4320	0.2453	0.5915	0.4511
CM3	0.0076	0.1024	0.2329	0.1611	0.2918	0.1111	0.3535	0.2177
CM4	0.0002	0.0004	0.0004	0.0013	0.0062	0.0221	0.0003	0.0003

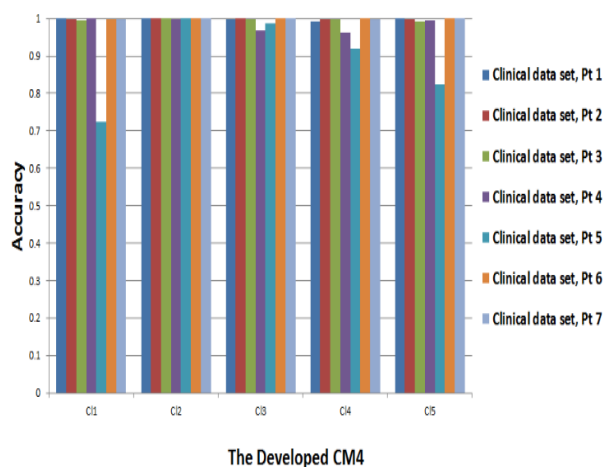


FIGURE 5. The accuracy obtained by the developed ensemble (CM4) for all five classes (C1–C5) in the clinical datasets.

approach in identifying different classes in the datasets, assisting the radiation oncologist who is handling the PET volumes in the clinical diagnosis of squamous cell carcinoma patients. The developed system was implemented and tested on a PC with a 3GHz processor and 16GB RAM, running a 64-bit Windows 10 operating system, x 64-based processor. The computational time for the ensemble model was 0.8 s.

Fig. 6 shows the significant improvement in accuracy achieved using CM4 for classifying and detecting the region of interest (tumor). An improvement of 100% was achieved for class 5 (tumor), as FFNN was not able to detect any voxel of the tumor in any of the seven patients, while CM4 detected the tumor in all patients, with an average accuracy above 98%. CM4 outperformed all eight other approaches in detecting all the recommended classes, particularly the target class (tumor).

On the other hand, for Pt 7, 1 voxel related to class 1 was misclassified into class 3 and 4 voxels related to class 4 was misclassified into class 1. Table 4 presents the MCV generated for the all patient datasets, where the lowest MCV

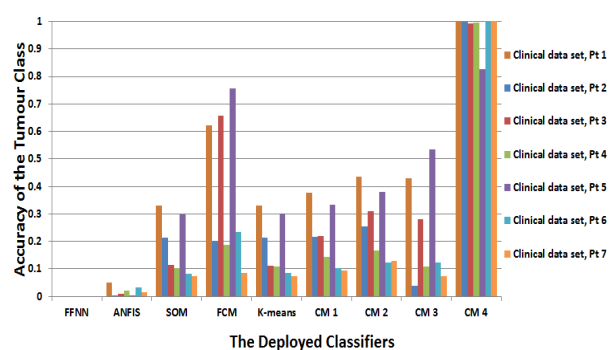


FIGURE 6. A comparison between the new ensemble (CM4) and other approaches (FFNN, ANFIS, SOM, FCM, K-means, CM1, CM2 and CM3) based on the Acc for tumor detection (class 5).

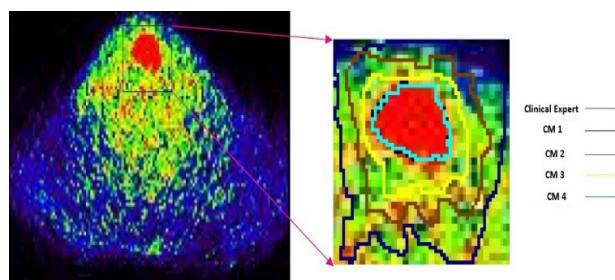


FIGURE 7. Representative segmentation results for the clinical dataset from Pt 1, where the black boundary represents the clinical expert estimation. The light blue boundary almost exactly overlaps the clinical expert boundary.

value of 0.0003 was obtained in datasets from Pt 6 and Pt 7. The MCV values for all classifiers are presented in Table 4. Fig. 7 shows representative segmentation results for the clinical dataset from Pt 1, where the black boundary represents the clinical expert estimation. The best match with this boundary is achieved by CM4, where the light blue boundary is almost overlapping the clinical expert boundary.

V. DISCUSSION

To evaluate the performance of CM4 in general for all classes, not just the region of interest, an average accuracy (AAcc)

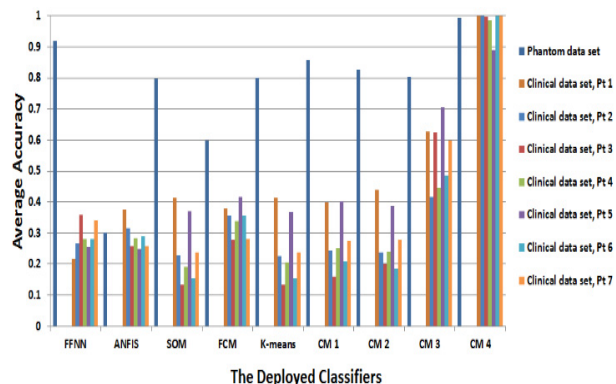


FIGURE 8. A comparison between the new ensemble (CM4) and the remaining approaches (FFNN, ANFIS, SOM, FCM, K-means, CM1, CM2 and CM3) based on the AAcc.

metric was introduced. This metric considers the average of the accuracies of all the subject classes. This value provides a general assessment for each of the classifiers considered. The AAcc for the experimental phantom dataset achieved by CM4 (99.39%) was the best among all the classifiers, CM1, CM2 and CM3. The AAcc obtained by CM4 for Pt 1's dataset was 99.83%, representing a significant improvement level of 78% in comparison to the lowest accuracy of 21.71% achieved through the FFNN approach. The obtained AAccs of CM4 were 99.94% and 99.93% for Pt 6 and Pt 7, respectively. The AAcc achieved for these two datasets are higher than that obtained for Pt 1, Pt 2, Pt 3, Pt 4 or Pt 5. This robust result indicates that the developed system has a stabilized robust performance with a higher accuracy than the other approaches.

Fig. 8 illustrates a complete comparison of the accuracy of the new ensemble (CM4) and the remaining proposed approaches for all the analyzed phantom and clinical datasets (Pts 1–7). In addition to the previously discussed objective performance evaluations for the achieved results by the new system, a comprehensive subjective evaluation in light of the clinician's expertise has been carried out to validate the performance of this approach.

VI. CONCLUSION

This study proposed an efficient PET volume handling approach for a robust PET volume analysis of squamous cell carcinoma. This approach was based on FFNN, ANFIS, SOM, FCM, and K-means. After the initial evaluation of these five classifiers, three ensemble classifiers (CM1, CM2, and CM3) were built using different methodologies such as weighted averaging, voting and averaging techniques. As the performance evaluation of these three ensemble classifiers did not reveal a significant level of accuracy for classifying the region of interest (especially for the clinical datasets), an optimized novel approach (CM4) based on the combination of a modified particle swarm optimization (PSO) and weighted averaging was developed for PET volume analysis. This approach overcame the misclassification problem associated with the previous approaches (CM1, CM2 and CM3). All the initial and developed approaches were evaluated

using experimental NEMA IEC body phantom and clinical PET studies for laryngeal squamous cell carcinoma patients. Superior results were obtained through the new optimized ensemble/committee machine (CM4) when compared to the results from the other approaches and the non-optimized ensembles. The proposed approach can identify the region of interest (tumor) accurately and precisely. The average accuracy obtained for the clinical studies of all patients (Pts 1–7) is 98.11%.

Promising results were achieved, for the clinical datasets. Regarding the NEMA body phantom dataset, the proposed approach achieved an overall accuracy of 99.39%. This accuracy was the highest in comparison with the accuracy of the other approaches and the non-optimized ensembles. The best improvement achieved by the ensemble CM4 was for the patients' datasets, where the overall voxels of the region of interest (tumor), were accurately allocated to the right class.

Different performance metrics were employed to validate the achieved performance. For example, an MCV value of 0.0002 was achieved for the NEMA phantom, and an MCV of 0.0003 was achieved for the pharyngolaryngeal clinical studies. The promising results achieved using different types of PET data, in particular for clinical data, indicated the stability and robustness of the proposed approach. Achieving an average accuracy of around 98% and matching the gold standard results fulfilled the aim of this paper, which was to assist clinicians in accurately and precisely analyzing the significant volumes of PET images.

ACKNOWLEDGMENT

The authors would like to thank John Lee of the Centre of Molecular Imaging, Radiotherapy and Oncology at UCL for the medical data support. This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, under Grant DF 3196111441.

REFERENCES

- [1] I. Ketata, L. Sallemi, F. Morain-Nicolier, M. Ben Slima, A. Cochet, K. Chtourou, S. Ruan, and A. Ben Hamida, "Factor analysis-based approach for early uptake automatic quantification of breast cancer by ^{18}F -FDG PET images sequence," *Biomed. Signal Process. Control*, vol. 9, pp. 19–31, Jan. 2014.
- [2] S. Vauclin, C. Michel, I. Buvat, K. Doyeux, A. Edet-Sanson, P. Vera, I. Gardin, and S. Hapley, "Monte-Carlo simulations of clinically realistic respiratory gated ^{18}F -FDG PET: Application to lesion detectability and volume measurements," *Comput. Methods Programs Biomed.*, vol. 118, no. 1, pp. 84–93, 2015.
- [3] G. Roman-Jimenez, R. De Crevoisier, J. Leseur, A. Devillers, J. D. Ospina, A. Simon, P. Terve, and O. Acosta, "Detection of bladder metabolic artifacts in ^{18}F -FDG PET imaging," *Comput. Biol. Med.*, vol. 71, pp. 77–85, Apr. 2016.
- [4] J. M. Mateos-Pérez, M. L. Soto-Montenegro, S. Peña-Zalvidea, M. Desco, and J. J. Vaquero, "Functional segmentation of dynamic PET studies: Open source implementation and validation of a leader-follower-based algorithm," *Comput. Biol. Med.*, vol. 69, pp. 181–188, Feb. 2016.
- [5] M. Aristophanous, B. C. Penney, and C. A. Pelizzari, "The development and testing of a digital PET phantom for the evaluation of tumor volume segmentation techniques," *Med. Phys.*, vol. 35, no. 7Part1, pp. 3331–3342, Jun. 2008.
- [6] M. S. Sharif and A. Amira, "An intelligent system for pet tumour detection and quantification," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2625–2628.

- [7] T. Doshi, J. Soraghan, L. Petropoulakis, G. Di Caterina, D. Grose, K. MacKenzie, and C. Wilson, "Automatic pharynx and larynx cancer segmentation framework (PLCSF) on contrast enhanced MR images," *Biomed. Signal Process. Control*, vol. 33, pp. 178–188, Mar. 2017.
- [8] K. Ballangan, X. Wang, M. Fulham, S. Eberl, and D. D. Feng, "Lung tumor segmentation in PET images using graph cuts," *Comput. Methods Programs Biomed.*, vol. 109, no. 3, pp. 260–268, 2013.
- [9] Y. Li, E. Porter, A. Santorelli, M. Popović, and M. K. Coates, "Microwave breast cancer detection via cost-sensitive ensemble classifiers: Phantom and patient investigation," *Biomed. Signal Process. Control*, vol. 31, pp. 366–376, Jan. 2017.
- [10] M. S. Sharif, M. Abbod, and A. Amira, "Neuro-fuzzy based approach for analysing 3D PET volume," in *Proc. IEEE Int. Conf. Develop. eSyst. Eng. (DeSE)-Special Session, Intell. Techn. Cancer Res.*, Dec. 2011, pp. 158–163.
- [11] G. P. Zhang, "Neural networks for classification: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 30, no. 4, pp. 451–462, Nov. 2000.
- [12] F. Roli and J. Kittler, "Multiple classifier systems," in *Proc. 3rd Int. Workshop MCS, Cagliari, Italy*, vol. 2364, 2002, pp. 27–41.
- [13] R. Duin, "The combining classifier: To train or not to train?" in *Proc. Int. Conf. Pattern Recognit.*, Quebec City, QC, Canada, Aug. 2002, pp. 765–770.
- [14] T. Windeatt and F. Roli, Eds., "Multiple classifier systems," in *Proc. 4th Int. Workshop, MCS, Guilford, U.K.*, vol. 2709, 2003, pp. 106–114.
- [15] M. S. Sharif, M. Alsallal, and L. Herghelegiu, "An effective TeleHealth assistive system to support senior citizen at home or care-homes," in *Proc. Int. Conf. Comput., Electron. Commun. Eng. (iCCECE)*, Aug. 2018, pp. 113–117.
- [16] J. Kittler and F. Roli, "Multiple classifier systems," in *Proc. 2nd Int. Workshop MCS, Cambridge, U.K.*, vol. 2096, 2001, pp. 136–147.
- [17] H. Kim, H. Kim, H. Moon, and H. Ahn, "A weight-adjusted voting algorithm for ensembles of classifiers," *J. Korean Stat. Soc.*, vol. 40, no. 4, pp. 437–449, Dec. 2011.
- [18] M. Termenon and M. Graña, "A two stage sequential ensemble applied to the classification of Alzheimer's disease based on MRI features," *Neural Process. Lett.*, vol. 35, no. 1, pp. 1–12, Oct. 2011.
- [19] G. Giacinto and F. Roli, "Design of effective neural network ensembles for image classification purposes," *Image Vis. Comput.*, vol. 19, nos. 9–10, pp. 699–707, Aug. 2001.
- [20] X. B. Song, Y. Abu-Mostafa, J. Sill, H. Kasdan, and M. Pavel, "Robust image recognition by fusion of contextual information," *Inf. Fusion*, vol. 3, no. 4, pp. 277–287, Dec. 2002.
- [21] F. Bovolo, L. Bruzzone, and L. Carlini, "A novel technique for subpixel image classification based on support vector machine," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2983–2999, Nov. 2010.
- [22] M. Shi, A. Bermak, S. B. Belhouari, and P. C. H. Chan, "Gas identification based on committee machine for microelectronic gas sensor," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 5, pp. 1786–1793, Oct. 2006.
- [23] D. Morrison and L. C. De Silva, "Voting ensembles for spoken affect classification," *J. Netw. Comput. Appl.*, vol. 30, no. 4, pp. 1356–1365, Nov. 2007.
- [24] X. Mu, J. Lu, P. Watta, and M. H. Hassoun, "Weighted voting-based ensemble classifiers with application to human face recognition and voice recognition," in *Proc. Int. Joint Conf. Neural Netw.*, Jun. 2009, pp. 2168–2171.
- [25] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1885–1896, Aug. 2009.
- [26] F. Veredas, H. Mesa, and L. Morente, "Binary tissue classification on wound images with neural networks and Bayesian classifiers," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 410–427, Feb. 2010.
- [27] L. A. da Silva, "Classification of breast masses using a committee machine of artificial neural networks," *J. Electron. Imag.*, vol. 17, no. 1, Jan. 2008, Art. no. 013017.
- [28] L. Wei, Y. Yang, R. M. Nishikawa, and Y. Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Trans. Med. Imag.*, vol. 24, no. 3, pp. 371–380, Mar. 2005.
- [29] B. Liu, Q. Cui, T. Jiang, and S. Ma, "A combinational feature selection and ensemble neural network method for classification of gene expression data," *BMC Bioinf.*, vol. 5, no. 1, p. 136, 2004.
- [30] Y. Chen, F. Chen, J. Y. Yang, and M. Q. Yang, "Ensemble voting system for multiclass protein fold recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 22, no. 04, pp. 747–763, Nov. 2011.
- [31] C. Ji and S. Ma, "Combinations of weak classifiers," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 32–42, Jan. 1997.
- [32] D. M. J. Tax, M. V. Breukelen, R. P. W. Duin, and J. Kittler, "Combining multiple classifiers by averaging or by multiplying?" *Pattern Recognit.*, vol. 33, no. 9, pp. 1475–1485, Sep. 2000.
- [33] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 281–286, Feb. 2002.
- [34] M. Clerc, "Particle swarm optimization," ISTE, London, U.K., Tech. Rep. QC20.7.M27C5513, 2006.
- [35] M. Clerc and J. Kennedy, "The particle swarm—explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, Aug. 2002.
- [36] Z. Bao, F. Wang, X. Zhao, and T. Watanabe, "Fault-tolerant image filter design using particle swarm optimization," *Artif. Life Robot.*, vol. 16, no. 3, pp. 333–337, Dec. 2011.
- [37] J. Robinson, S. Sinton, and Y. Rahmat-Samii, "Particle swarm, genetic algorithm, and their hybrids: Optimization of a profiled corrugated horn antenna," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, vol. 1, Jun. 2002, pp. 314–317.
- [38] J. Kennedy and W. M. Spears, "Matching algorithms to problems: An experimental test of the particle swarm and some genetic algorithms on the multimodal problem generator," in *Proc. IEEE Int. Conf. Evol. Comput. IEEE World Congr. Comput. Intell.*, May 1998, pp. 78–83.
- [39] C. Jonsson, R. Odh, P.-O. Schnell, and S. A. Larsson, "A comparison of the imaging properties of a 3- and 4-ring biograph PET scanner using a novel extended NEMA phantom," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, Oct./Nov. 2007, pp. 2865–2867.
- [40] M. D. R. Thomas, D. L. Bailey, and L. Livieratos, "A dual modality approach to quantitative quality control in emission tomography," *Phys. Med. Biol.*, vol. 50, no. 15, pp. N187–N194, Jul. 2005.
- [41] H. Bergmann, G. Dobrozemsky, G. Minear, R. Nicoletti, and M. Samal, "An inter-laboratory comparison study of image quality of PET scanners using the NEMA NU 2-2001 procedure for assessment of image quality," *Phys. Med. Biol.*, vol. 50, no. 10, pp. 2193–2207, Apr. 2005.
- [42] J. M. Wilson and T. G. Turkington, "Multisphere phantom and analysis algorithm for PET image quality assessment," *Phys. Med. Biol.*, vol. 53, no. 12, pp. 3267–3278, May 2008.
- [43] Data Spectrum Corporation. (2017). *NEMA IEC Phantom*. [Online]. Available: <http://www.spect.com/products-pet.html>
- [44] J.-F. Daisne, M. Sibomana, A. Bol, G. Cosnard, M. Lonneux, and V. Grégoire, "Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: Accuracy, reproducibility and consistency," *Radiotherapy Oncol.*, vol. 69, no. 3, pp. 237–245, Dec. 2003.
- [45] J.-F. Daisne, T. Duprez, B. Weynand, M. Lonneux, M. Hamoir, H. Reyckler, and V. Grégoire, "Tumor volume in pharyngolaryngeal squamous cell carcinoma: Comparison at CT, MR imaging, and FDG PET and validation with surgical specimen," *Radiology*, vol. 233, no. 1, pp. 93–100, Oct. 2004.
- [46] X. Geets, J. A. Lee, A. Bol, M. Lonneux, and V. Grégoire, "A gradient-based method for segmenting FDG-PET images: Methodology and validation," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 34, no. 9, pp. 1427–1438, Mar. 2007.
- [47] R. C. Eberhart and Y. Shi, "Particle swarm optimization: Developments, applications and resources," in *Proc. Congr. Evol. Comput.*, vol. 1, May 2001, pp. 81–86.
- [48] A. Carlisle and G. Doizier, "An off-the-shelf PSO," in *Proc. Part. Swarm Optim. Workshop*, 2001.
- [49] T. Kohonen, *Self-Organizing Maps*. Spain: Springer, 2001.
- [50] S. Miyamoto, H. Ichihashi, and K. Honda, *Algorithms for Fuzzy Clustering: Methods in C-Means Clustering With Applications*. Spain: Springer, 2008.
- [51] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Upper Saddle River, NJ, USA: Prentice-Hall, 1988.
- [52] S. H. Ebenuwa, M. S. Sharif, M. Alazab, and A. Al-Nemrat, "Variance ranking attributes selection techniques for binary classification problem in imbalance data," *IEEE Access*, vol. 7, pp. 24649–24666, 2019.
- [53] D. L. Weakliem, "A critique of the Bayesian information criterion for model selection," *Sociolo. Methods Res.*, vol. 27, no. 3, pp. 359–397, 1999.
- [54] C. T. Volinsky and A. E. Raftery, "Bayesian information criterion for censored survival models," *Biometrics*, vol. 56, no. 1, pp. 256–262, May 2004.
- [55] J. Strzelczyk, "The essential physics of medical imaging," *Health Phys.*, vol. 85, no. 2, p. 242, Aug. 2003.



MHD SAEED SHARIF is currently an Associate Professor with the School of Architecture, Computing and Engineering, UEL. He is also the Program Leader for the master's in computer science and also led the research development of research projects associated with different industries. His research interests include artificial intelligence, innovative tele-health, medical technology, digital health care, medical assistive technology, medical image analysis and visualization, intelligent diagnosis systems, big data, medical biotechnology, smart biomedical image, and bio signal acquisition (MRI, PET, and EEG). He is working closely with clinicians and policy makers nationally and internationally to improve the clinical settings and the healthcare systems.

Dr. Saeed is a Fellow of the U.K. Higher Education Academy. He received many academic and research awards. He has participated in many national and international conferences, e.g., ICIP. He has served as a Reviewer for many journals, e.g., IET IPJ and Elsevier CBMJ.



MAYSAM ABBOD received the Ph.D. degree in control engineering from The University of Sheffield, U.K., in 1992. He is currently a Reader of electronic systems with the Department of Electronic and Computer Engineering, Brunel University London, U.K. He has authored more than 50 articles in journals, 9 chapters in edited books, and more than 50 papers in refereed conferences. His current research interests include intelligent systems for modeling and optimization. He is a member of the IET, U.K., and a Chartered Engineer in U.K. He is serving as an Associate Editor for the *Engineering Application of Artificial Intelligence* (Elsevier).



ALI AL-BAYATTI received the Ph.D. degree in computer science in 2009. He worked with leading organizations such as Deloitte, Airbus, Elektrotbit Automotive and Rolls-Royce, among others. He is currently an Associate Professor with De Montfort University. He is also the Subject Leader of cyber security with the Cyber Technology Institute. He is also a Visiting Professor at multiple institutes and a member of the Oman research council. He is one of the main factors behind a generated annual income of £1.2 Million at the Cyber Technology Institute. His current research is multidisciplinary it includes vehicular ad hoc networks, driver behavior, cyber security, and smart technologies that promote collective intelligence and its applications range from promoting comfort to enabling safety in critical scenarios. He serves on multiple Editorial Boards and, is on the Scientific Advisory Boards of multiple institutes in Gulf and Europe.



ABBES AMIRA (Senior Member, IEEE) received the Ph.D. degree in computer engineering from Queen's University Belfast, U.K., in 2001. Since 2001, he has been taking many academic and consultancy positions in U.K., Europe, Asia, and Middle East. In U.K., he has previously taken academic and leadership positions at Queen's University Belfast, Brunel University London, and the University of Ulster. From 2017 to 2019, he was an Associate Dean for research and graduate studies with the College of Engineering, Qatar University, Qatar. He has taken visiting professor positions at the University of Tun Hussein Onn, Malaysia, and the University of Nancy, Henri Poincare, France. He is currently an Associate Dean for research and innovation with the Faculty of Computing, Engineering and Media, De Montfort University, U.K. During his career to date, he has been successful in securing substantial funding from government agencies and industry. He has supervised more than 25 Ph.D. students and has more than 350 publications in top journals and conferences in the area

of embedded systems, the IoT, image, and signal processing. He has also conducted consultancy services for several government agencies and companies in the private sector. His research interests include embedded systems, high-performance computing, big data and the IoT, connected health, image and vision systems, biometric and security.

Prof. Amira has participated as a member and a Guest Editor of the editorial board in many international journals including recent special issues in the IEEE IoT JOURNAL and *Pattern Recognition* (Elsevier). He is a Fellow of IET and of the Higher Education Academy, and a Senior Member of ACM. He received many international awards, including the 2008 VARIAN prize offered by the Swiss Society of Radiobiology and Medical Physics, the CAST Award, the IET Premium Award, in 2017, the DELL-EM Envision the Future, in 2018, and many best paper and recognition awards in the IEEE international conferences and events.



AHMED S. ALFAKEEH received the degree in computer science from King Abdulaziz University, in 2005, the M.Sc. degree in IT from De Montfort University, in 2010, and the Ph.D. degree from the Software Technology Research Laboratory, a research institute established within, De Montfort University, Leicester, U.K., in 2015. He is currently an Assistant Professor with the Information System Department, Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, Saudi Arabia. He is also a Vice Dean of the Research and Consultation Institute for Development. His research projects involve feature interactions detection and resolution in context-aware systems, smart homes systems, software engineering, the IoT, and data science.



BAL SANGHERA received the Ph.D. degree in medical physics imaging from the Institute of Cancer Research combined with Imperial College. He was involved in physics locum and held postdoctoral positions at Hammersmith Hospital and Imperial College in imaging, robotics, and biomechanics. He is currently a PET-CT Physicist, a State Registered Clinical Scientist, and a Medical Physics Expert with the Paul Strickland Scanner Centre, Mount Vernon Hospital, London. He is also a Medical Physics Examiner with the Royal College of Radiology. His typical routine as a medical physicist was duties with strong research commitments. He supervised M.Sc. and Ph.D. students and current one is involved with textural analysis in PET-CT and MRI of the prostate. He is also supervising a Ph.D. student at Birmingham University in PET-CT in head and neck cancer. He currently has 36 peer-reviewed accepted publications to date. He is currently publishing a national PET dose calibrator study looking at device differences in 13 PET-CT centres using a Ge-68 sealed source syringe as a surrogate for F-18.

He was a Trustee of the Institute of Physics and Engineering (IPEM), IPEM Honorary Secretary, and a member of the Business and Finance and Committee. He is also an Inaugural Chief Scientific Officer's Knowledge Transfer Partnership Associate with NHS England heading a PET-CT harmonization national study using a Ge-68 NEMA image quality phantom.

...