Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 4, August 2026, pp. 3314∼3327
ISSN: 2302-9285, DOI: 10.11591/eei.v15i4.11491

Improved chicken swarm optimization with RNNs for multi-label classification in imbalanced data

M. Priyadharshini[1], Narendruni Lakshmi Priya[2], Anitha Vippdapu[1], Subrata Chowdhury[3], Thi-Thu Nguyen[4], Duc-Tan Tran[5], Duc-Nghia Tran[6]

[1]Department of Computer Science and Engineering, Faculty of Science and Technology (IcfaiTech), ICFAI Foundation for Higher Education, Hyderabad, India
[2]Department of Computer Science and Engineering, Mahatma Jyothiba Phule Telangana Backward Classes Welfare Residential Degree College (Women), Suryapet, India
[3]Department of Master of Computing Application, C. Abdul Hakeem College of Engineering and Technology Melvisharam, Ranipet, India
[4]Faculty of Embedded Systems and Integrated Circuits, School of Electrical and Electronic Engineering, Hanoi University of Industry, Hanoi, Vietnam
[5]Department of Electronics and Communication, Faculty of Electrical and Electronic Engineering, Phenikaa University, Hanoi, Vietnam
[6]Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

Article Info

Article history:

Received Sep 26, 2025
Revised Jul 6, 2026
Accepted Jul 14, 2026

Keywords:

Adaptive synthetic data
Improved chicken swam optimization
Multi-label classification
Principle component analysis
Recurrent neural network

ABSTRACT

Multi-label classification (MLC) seeks to make multiple predictions for an instance by identifying associations between labels, thereby enhancing prediction capability. The paper presents a new method for combining improved chicken swarm optimization (ICSO) feature selection with recurrent neural networks (RNNs) for MLC. ICSO overcomes feature selection issues and minimizes noise and redundancy in imbalanced datasets, whereas RNNs learn label dependencies to achieve higher accuracy. It is initially normalized using the Z-score method and then analyzed for dimensionality reduction using principal component analysis (PCA). Our approach is more accurate, more precise, better at recall, and achieves higher F-measure and specificity, and a lower error rate than current methods such as multi-label k-nearest neighbors (ML-KNN), fuzzy rough set learning with label-specific features (FRS-LIFT), and adaptive synthetic data for multi-label classification (ASD-MLC). The paper shows that ICSO is useful for augmenting RNN-based multi-label classification and may be applied in medical diagnosis and bioinformatics.

Corresponding Author:

Duc-Tan Tran
Department of Electronics and Communication, Faculty of Electrical and Electronic Engineering
Phenikaa University
Hanoi, Vietnam
Email: tan.tranduc@phenikaa-uni.edu.vn

1. INTRODUCTION

Instance (or) attribute can be connected to numerous labels called multi-label classification in supervised learning. In contrast, each occurrence in a conventional single-label classification jobs (i.e., binaries or multi-classes) get connected singular class labels. Multi-label contexts are gaining popularity which are being applied on a variety of domains including texts, audio, images, videos, bioinformatics, and their referrals [1]. The most popular method for classifying data with multiple labels are training different classifiers on labels.

The binary relevance (BR) transformation is the standard name for this process. In essence, an issue with many labels is reduced to a single binary problem for each label, and any commercial binary classifier is then applied to each of these problems separately. The fact that relationships between labels are not explicitly modeled and those strategies to account for these dependencies essentially limit entire multi-label methods [2].

The so-called issue transformation methods (where multi-label issues are turned into multi-class or binary issues) have been used to create numerous effective multi-label algorithms up to this point. These techniques repeatedly copy or iterate across the feature space in memory. The majority of the best performing techniques also include ensembles, such as with boosting, decision trees, and support vector machines (SVMs). In other words, majority of competing approaches might greatly profit. MLC issues belong into one of two groups: algorithm adaptation techniques or problem transformation. The label powerset (LP) problem transformation method considers each distinct collection of labels for a multi-label dataset as a single label.

Traditional methods issues for multi-label classification: The efficacy of deep learning (DL) models on the physical examination data is assessed utilizing a variety of machine learning (ML) techniques [3], [4]. A series of conventional ML techniques are applied in this field, along with a few techniques created especially to address multi-label classification issues [5], [6]. While the other conventional approaches applied to multi-label issues, they typically require some dataset manipulation for the technique to accurately comprehend dataset targets. Stated differently, it converts a dataset with multiple labels into a single-label dataset that has several classifications. This kind of conversion has been handled utilizing a wide variety of methods. Multi-label classification issues fall into two broad categories: problem transformation and algorithmic adaptation techniques. An approach to problem transformation is referred to as the LP, in which every distinct set of labels for a dataset with multiple labels is regarded as a single label. It is thought of as a power set, this particular set of labels. These power sets are utilized for training a classifier so that it can provide predictions. This specific strategy is employed in several of the ways that follow to handle multi-label categorization. Nevertheless, modifying the data to fit this format has certain disadvantages. LP datasets frequently result in a large number of described classes and few training sample instances for each class. DL approaches have an added benefit over comparable issue transformation methods in that they can train on the original data without requiring any kind of data conversion. These DL techniques are similar to those employed for system adaptation.

A multi-label problem's early investment in lowering the number of feature variables is far more likely to produce appreciable speed-ups during learning and classification since more succinct representations of the feature space are often significantly more useful than in the single-label situation. However, this strategy has only been taken into account in a small amount of the multi-label literature. The implicit assumption when building a model from the raw instance data is that the labels come from this data and can be retrieved straight from it. In most cases, however, certain abstract notions serve as the basis for both the labels and the feature variables. To avoid these problems in proposed work introduced an efficient model for multi label classification.

- Initially the Z-score; normalizations are executed on inputs;
- Principal component analysis (PCA) is employed for dimensionality reduction;
- Feature selection is done by using improved chicken swarm optimization (ICSO) to reduce unwanted and noise attributes;
- This work uses recurrent neural networks (RNNs) for multi label classifications in this work.

The paper proposes an improved version of the ICSO feature selection algorithm and RNNs MLC algorithm to the problem of imbalanced datasets that is not successfully addressed by the existing methods.

2. RELATED WORKS

MLC is an important task in ML, where each instance is associated with multiple labels. Traditional methods, such as BR and problem transformation techniques like LP, transform MLC problems into binary or multi-class classification problems. The approaches, however, do not technically elaborate the label dependencies and the issue of unbalanced datasets; crucial limitations in most real-world applications [1], [2].

In MLC, feature selection is a very important process particularly with high-dimensional data. Techniques of feature selection such as minimum redundancy maximum relevance (mRMR) have been extensively used and optimization methods such as the ant colony optimization (ACO) and genetic algorithms (GA) have been utilized to help minimize the dimension size and enhance the results of classification. MLC have also been done by the use of chicken swarm optimization (CSO), a biologically-inspired optimization algorithm, in feature selection. But CSO is susceptible to local optima problems, which may constrain its usefulness. Re-

cent progress brought about the emergence of ICSO, which overcomes these concerns by evolving a mutation operator, which has a better capacity to navigate the feature space better and pick up the features to be used in classification tasks [3], [4]. The work is based on ICSO, which is used to select features efficiently to classify multiple labels, in particular, imbalanced datasets.

MLC has recently received a lot of interest because of the dependencies between labels with time using DL, especially RNNs, which can capture label dependencies across time. RNNs variants such as long short-term memory (LSTM) and gated recurrent units (GRU) have proven incredibly effective in MLC challenges and especially in data that has temporal or sequential dependencies. Applications of these models have been effective in tackling partly text classification, image classification and bioinformatics problems which also represent multi-label problems. Recent research investigated collaborative learning techniques to acquire label semantics and label-specific features with DL models to multi-label learning tasks [5], [6]. Nevertheless, these models are usually fraught with difficulties encountered when using the imbalanced datasets, which is where our method of using ICSO in the selection of features and RNN in the classification becomes useful.

A combination of optimization algorithms and ML models to create hybrid approaches has proven effective in enhancing MLC. A modified ACO to feature selection has shown superior classification and less run time. Multi-target regression techniques have been used in other similar studies to reduce discrepancies between target variables. Our work presents a new hybrid methodology, with the ICSO feature selection method together with RNNs MLC method, overcoming the problem of label dependence and imbalanced dataset better than their respective counterparts [7], [8].

The problem of imbalanced datasets is a significant issue in MLC since the models are frequently biased to one of the most common classes, and thus fail to predict the labels of the minority classes. A number of studies have used synthetic data generation methods, including synthetic minority over-sampling technique (SMOTE) and adaptive synthetic sampling to cope with this issue. These measures assist in creating a trade-off between the data, by creating artificial examples of the minority group. ICSO will scale down irrelevant features in our method, and Imbalanced data is dealt with through the use of RNNs to model the label dependencies. The performance of our method is better in accuracy, precision, recall and specificity compared to existing methods [9], [10].

Recent developments in the feature selection methods such as gray wolf optimization (GWO) and ACO have demonstrated a significant degree of enhancement in classifications. The techniques are commonly applied together with ML models to enhance the effectiveness of multi-label classification. The combination of ICSO and RNNs in our solution does not only optimize the feature selection process, the problems of unequal datasets are solved as well and this makes our method more resistant and more precise than the traditional methods [11].

3. PROPOSED METHOD

Proposed work steps are: i) preprocessing based on Z-score normalizations, the initial components of the proposed model, ii) PCA reduces dimensionalities, iii) feature selections are done using ISCO, iv) RNNs finally classify multi-labels, and v) performance comparison done with assessment metrics including accuracy, recall, precision, f-measure, error rate and specificity. Figure 1 depicts overall methodology of the proposed scheme.

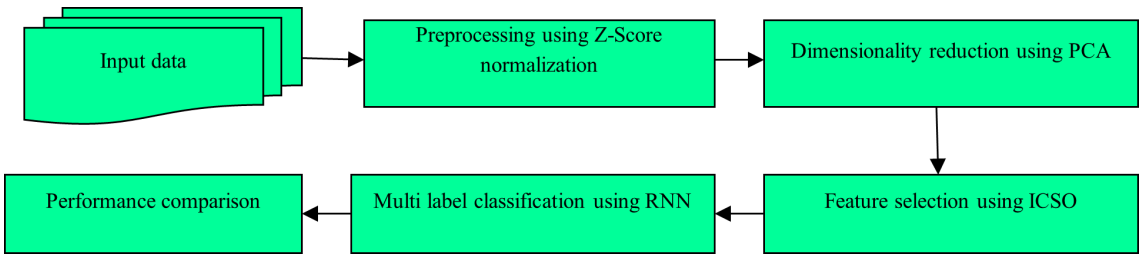

Figure 1. Overall methodology of the proposed model

3.1. Data normalization using Z-score normalization model

Data pre-processing is used to confirm the avoidance of data overwhelmed with each other. Data from various scales are converted to data on same scales using Z-score normalization. Values are normalized using means and standard deviations of features A in Z-score normalizations method and as depicted (1):

$$v' = \frac{v - \bar{A}}{\sigma_A} \tag{1}$$

Where, v' stands for new data entries, v represents old data entries, σ_A represents standard deviations of A, and $\bar{A}$ implies means of A.

3.2. Dimensionality reduction using principle component analysis

These normalized inputs are then sent to the dimensionality reduction procedure to lessen the computational complexity after normalization [9]. PCA is used in this study to reduce the number of dimensions. PCA, a dimensionality reduction technique, combines the initial attributes in a linear manner to produce new features. Instances of given datasets discovered in d-dimensional spaces are mapped using PCA to k-dimensional subspace where. The k newly generated dimensions that make up principle components aim for maximal variances while avoiding variation that have already been taken into account by all of its prior components. As a result, the first component accounts for the highest volatility, whereas each succeeding component accounts for less variation [10]. In (2) represents principal components.

$$PC_i = a_1X_1 + a_2X_2 + \ldots + a_dX_d \tag{2}$$

Where PC_i represent principle components 'i', X_j represents original features of j, and a_j implies $X_j's$ numerical coefficients. A linear link among two variables' direction and size are expressed statistically as correlation. The covariance matrix is produced in the context of PCA using correlation. This square matrix displays the pair wise correlations of all variables in the dataset. The diagonal elements of covariance matrices represent variables' variances, while off-diagonal components represent covariance between different combinations of variables. Correlation coefficients are standardized measures of correlations in the range of -1 to 1, can obtain strengths and directions of variables' linear links. The perfect positive and negative correlations are shown by correlation coefficients of 1 and -1, respectively. The lack of a linear relationship among the two variables is indicated by a correlation value of 0. The initial variables are combined linearly to produce the principle components in PCA, which maximise the variation explained by the data. The correlation matrix is used to determine principal components.

Given a dataset with n variables, correlation matrices C are nxn symmetric matrices having the following components $(x_1, x_2, ..., x_n)$ in (3):

$$C_{ij} \ = \ (sd(x_i) \ * \ sd(x_j)) \ / \ cov(x_i, \ x_j) \tag{3}$$

Where $cov(x_i, x_j)$ simply correlations between variables x_i and x_j and $sd(x_i, x_j)$ represent standard deviations of variable $sd(x_i)$ and $d(x_j)$, respectively. Correlation matrices C can be expressed as (4):

$$C \ = \frac{X^T X}{(n-1)\ (n-1)} \tag{4}$$

X^T and X are variables of correlation matrices.

In the context of the PCA technique, the term "orthogonality" to the formation of the principle components as being orthogonal to one another and its restrictions are written as (5):

$$a_{i1} \ * \ a_{j1} \ + \ a_{i2} \ * \ a_{j2} \ + \ ... \ + \ a_{ip} \ * \ a_{jp} \ = \ 0 \tag{5}$$

if $i \ \neq \ j$ for all i and j, it implies that any two loading vectors for dissimilar principal components have their dot product as zero, which denotes that the principal components are orthogonal to one another. The eigenvectors are used to compute the key elements of the data. Data eigenvectors of covariance matrices reveal considerable variability. These coordinates are then used to build new coordinates for data is representations.

The letters C stand for the covariance matrix. $v_1, v_2, ..., v_p$ in mathematics, whereas the associated eigen values are denoted by the letters 1, 2,..., p. The eigenvectors are computed so that (6) holds:

$$C\ v_i\ = \lambda_i\ v_i \tag{6}$$

This implies eigenvectors v_i produce associated eigen values λ_i as their own scalar multiples when multiplied by covariance matrices C. The calculations of principal components in data using PCA method depends critically on the covariance matrices which are pxp matrices and compute pairwise covariance between data elements. Given data matrices X with n observations over p variables, the correlation matrices C can be defined as (7):

$$C\ =\ \left(\frac{1}{n}\right)\ *\ X^T X \tag{7}$$

Where X^T implies transpositions of X. The off-diagonal sections of C represent the variables' covariances, whereas the diagonal components of C represent the variables' variances.

3.3. Feature selection using improved chicken swam optimization

Once the dimensionality reduction process is completed it sends to the feature selection phase. In this work improved select features.

3.3.1. Chicken swam optimization

CSO is a meta-heuristic optimisation technique with biological influences. The software simulates both the motions of individual chickens and hierarchical structures of chicken swarms. A chicken swarm is organised hierarchically, with each group consisting of a roosters and large counts of hens and chicks [11]. Different types of chickens are governed by various rules of motion. In chicken social interactions, hierarchy is essential [12], [13]. The hens and roosters in groups' edges are more submissive with domineering hens that stay closer to head roosters in flocks and where the stronger hens will rule over the less competent ones. Local optimum characteristics are a common pitfall for traditional CSO. This work utilises the CSO mutation operator to get around this issue. Flip bit mutation was employed in this work. This mutation operator flips the bits in the genome of choice. That is, the genomic bit is changed from 1 to 0 and vice versa if it is currently 1.

3.3.2. Improved chicken swarm optimization

Chicks' behaviours form the base for of ICSO's mathematical models and summarized as:

- Flocks of chicks have many groups with roosters in control followed by hens and chicks.
- Fitness values of chickens depict swarms' hierarchies where roosters function as group leaders, have good fitness values in contrast to chicks which have very low fitness values. The other group would be the chickens.
- A group's mother-child attachment, dominance relationship, and swarm hierarchy will not change and statuses get updated after time steps (G).
- The counts of virtual chickens in swarms are denoted by letters RN (roosters), CN (chicks), HN (hens), and MN (mother hens). Figure 2 shows the ICSO flow chart. Their locations are indicated by (8):

$$x_{i,j(i\in[i,.....N],\ j\in[1,.....,D])}, \tag{8}$$

Rooster movements: according to (9) and (10), roosters with higher fitness levels forage for food in multiple locations than those with lowered fitness.

$$x_{i,j}^{t+1} = x_{i,j}^{t} * \left(1 + Randn\left(0, \sigma^2\right)\right) \tag{9}$$

$$\sigma^2 \begin{cases} 1, & if f_i \leq f_k\ ,\ | \\ exp\left(\frac{f_k - f_i}{|f_i| + \in}\right) & otherwisek \in [1, N]\ ,\ k \neq i, \end{cases} \tag{10}$$

Where $x_{i,j}$ implies indices I of chosen roosters, $Randn(0, sigma^2)$ signify distributions (Gaussian) having zero as mean and standard deviation of σ^2 ,$\in$ s smallest constants assist in avoiding zero-division-errors, k stands for randomly selected roosters indices, f_i stands for $x_i's$ rooster values.

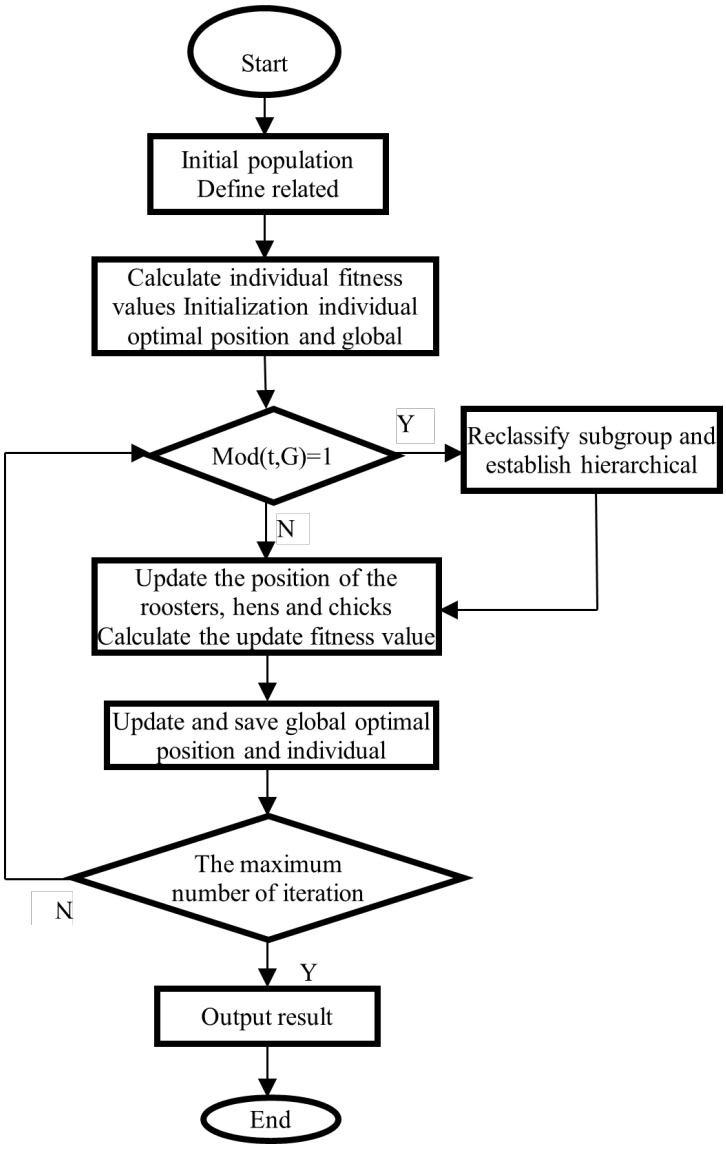

Figure 2. ICSO

Hen movements: these follow roosters for food, but when being confined they seize foods of other birds. More dominant hens have an edge over more submissive ones in competitions for food. In (12) and (13) are mathematical formulations of these events.

$$x_{(i,j)}^{(t+1)} = x_{(i,j)}^{t} + S_1\, rRand\left(x_{(r_{(1,j)}),}^{t} - x_{(i,j)}^{t}\right) + S_2\, Rand\left(x_{(r_{(2,j)}),}^{t} - x_{(i,j)}^{t}\right) \tag{11}$$

$$S1 = exp\left(\left(f_i - f_{r1}\right)/abs\left(f_i\right) + \in\right)) \tag{12}$$

$$S1 = exp\left(\left(f_{r_2-}\, f_i\right)\right) \tag{13}$$

Where Rand stands for consistent randomized number in the interval [0,1]. Roosters' indices are represented by $r_1 \in [1, \ldots.., N]$ i.e. ith hen group-mates, while $r_2 \in [1, \ldots.., N]$, suggests selected randomized indices of chickens (hens or roosters) from swarms.

Chick movement: the movement of chicks for food searches can be formulated as (14):

$$x_{i,j}^{t+1} = x_{i,j}^{t} + FL * \left(x_{m,j}^{t} - x_{i,j}^{t}\right) \tag{14}$$

Where $x_{i,j}^{t}$ stands for i^{th} chick's mothers; positions such that $m \in [1; N]$, FL stands for the speeds which chicks trail their mothers based on differences between chicks (FL) randomly selected in the range [0,2]. Feature values in ranges of 0 to 1 necessitates intelligent searches for locating best points in search spaces that maximize fitness functions. Given the training data, the CSO's fitness function aims to maximise classification accuracy across the validation set while retaining a minimal features counts, as stated in (15):

$$f_{\theta} = \omega * E + (1 - \omega) \frac{\sum_i \theta_i}{N} \tag{15}$$

Where f_{θ} implies θ vectors'fitness functions with 0/1 elements represent in gun selected features, N refers to counts of dataset features, E refers to error rates of classifiers and ω constants control classifier's features selected counts for better performances. The feature counts in the provided dataset match the utilized variable counts. Every variable has a range of [0, 1], and as the value of variables approach 1, associated features become a candidate for selections in classifications. The variables in determining individual fitness are thresholds, which set particular traits to be judged, as shown in (16):

$$f_{i,j} = \begin{cases} 1 \; if X_{i,j} > 0.5 \\ 0 \; otherwise, \end{cases} \tag{16}$$

Where X_{ij} refers to dimension values of search agents i at dimension j and update firefly locations; solutions, updated values at some dimensions might break constraints [0, 1], and therefore utilised as simple truncation rules to assure variable limits. The hierarchical swarm dynamics optimization (HSDO) algorithm is a population-based metaheuristic optimization technique inspired by the hierarchical behavior and social structure of a chicken swarm.

Algorithm 1. A HSDO using dynamic role-based position updates
1. Initialise RN, HN, CN, MN, G;
2. Initialise chickens in swarms at random.
3. X_i $(i = 1, 2, \ldots, N)$;
4. Set the max iterations count T_{max}.
5. while $T < T_{max}$ in iterations do
6. if $T\%G = 0$ then
7. Position fitness ratings of chickens (Ranks) and generate hierarchical orders in swarms.
8. Separate swarms into groups and identify connections between chicks and mother hens inside groups;
9. end
10. for chickens X_i in swarms do 11. if X_i represent rosters then
12. Update locations of X_i with (10);
13. end
14. if X_i represent hens then 15. Update locations of X_i with (1);
16. end
17. if X_i represent chicks then
18. Update locations of X_i with (15);
19. end
20. Use (16) to evaluate new answers;
21. If the new solutions are better than prior ones, update them.
22. end
23. end
24. Execute flip bit mutations on modified answers.
25. Evaluate new answers with (16);
26. end

3.4. Multi label lassification using recurrent neural network

Following feature selection, the chosen feature subsets are categorised using RNNs. We employ feed forward multi-layer perceptions as RNNs which use hidden layers (3) between inputs and outputs [14]. The goal of training RNNs is reproduce input data patterns much closer to output layers and n outputs correspond n training data features [15], [16]. Architecture of the RNN is shown in Figure 3. The number of hidden layers

empirically decrease average reconstruction errors across all training patterns. The weighed sums of inputs to layer k units I, represented as Iki, and activation functions Sk(Iki) decide on unit outputs.

$$\theta = I_{ki} = \sum_{j=0}^{L_{k-1}} w_{kij} Z_{(k-1)j} \tag{17}$$

Z_{kj} represents outputs from j^{th} units of k^{th} layers. L_k represents unit counts of k^{th} layer. The following represents activation functions of the hidden layers (k=2, 4):

$$S_k(\theta) = \tanh(a_k \theta) \qquad k = 2, 4 \tag{18}$$

Where a_k imply tuning variables and equal one. Activation functions of intermediate hidden layers (k=3) are similar to staircases, where N implies step counts or activation levels a_3 refers to controlling rates at which levels are advanced [17]:

$$S_3(\theta) = \frac{1}{2} + \frac{1}{2(k-1)} \sum_{j=1}^{N-1} \tanh\left[a_3 \left(\theta - \frac{j}{N}\right)\right] \tag{19}$$

Hidden activation levels are quantized into arrays with N discrete values: $0, \frac{1}{N1}, \frac{2}{N1}, \ldots 1$ [18]-[20]. Intermediate hidden layers' step-wise activations divide evenly distributed data points into groups of discrete vector values [21]-[25].

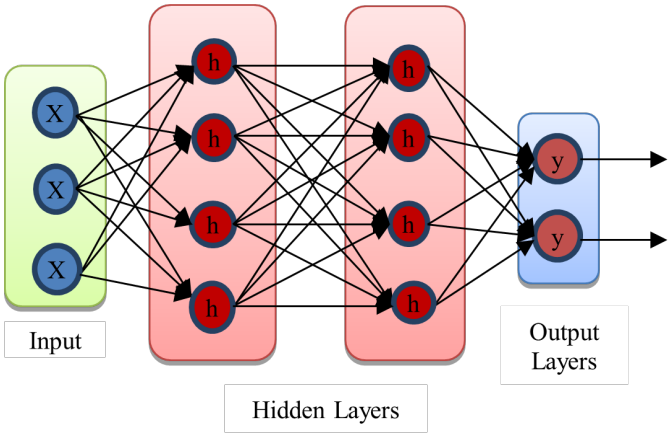

Figure 3. RNN

4. RESULTS AND DISCUSSION

4.1. Experimental procedure

This section discusses the recommended model's experimental results. Adaptive synthetic data for multi-label classification (ASD-MLC), feature selection using evolutionary algorithm for multi-label classification (FSEA-MLC), fuzzy rough set learning with label-specific features (FRS-LIFT), and the current multi-label k-nearest neighbors (ML-KNN) techniques are contrasted with the suggested RNN's values of precisions, recalls, accuracies, specificities, error rate, and f-measures. In this paper, Dermatology, Page Blocks, and Shuttle datasets [21] are collected from KEEL repository. These datasets were divided into training (70%) and testing (30%) datasets for each experiment by MATLABR2021(a) to assess the classification approaches (Table 1). The proposed approach is contrasted with other popular techniques including ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC algorithms. Metrics including accuracy, precision, recall, F-measure, specificity, and error rate are used to evaluate classifier performance.

The paper proposes an improved version of the ICSO feature selection algorithm and RNNs multi-label classification algorithm to the problem of imbalanced datasets that is not successfully addressed by the existing methods.

Table 1. Details of dataset

Dataset	Imbalanced ratio	Instance	Features	Labels
Dermatology	5.55	366	34	6
Page blocks	164	548	10	5
Shuttle	853	2175	9	7

4.2. Performance metrics

To rigorously assess the effectiveness of the proposed model in handling multi-label classification on imbalanced datasets, a comprehensive set of evaluation metrics is employed. These metrics—including precision, recall, accuracy, F-measure, error rate, and specificity—provide a multi-dimensional perspective on classification performance by capturing both positive and negative prediction behavior.

– Precisions is refers to percentages of results which are relevant and defined as (20):

$$Precision = \frac{true\ positive}{true\ positive + false\ positive} \tag{20}$$

– Recalls refer to percentages of total accurate relevant results classified by proposed algorithms and defined as (21):

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \tag{21}$$

– Accuracies are fractions of models' correct predictions and defined as (22):

$$Accuracy = \frac{tru\ positive + true\ negative}{Total} \tag{22}$$

– An F-scores are harmonic means of computed precisions and recalls as (23):

$$F\text{-}score = 2 \times \left[\frac{precision \times recall}{precision + recall}\right] \tag{23}$$

– Error rates refer to measuring degrees of prediction errors of models with respect to true models and computed as (24):

$$Error\ rate\ =\ 100 -\ accuracy \tag{24}$$

– Specificity: the percentage of true negatives that the algorithm accurately recognizes is known as specificity. And which is calculated as (25):

$$Specificity = \frac{true\ negatives}{true\ negative + false\ positives} \tag{25}$$

Figure 4 demonstrates the performance of the proposed RNN-based multi-label classification model in terms of accuracy in comparison to other existing algorithms, including ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC. The result of accuracy shows that the proposed RNN predictor performs much higher than the other models (including ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC), having the accuracy of 94.5789% on the Shuttle dataset.

Its recurrent structure allows the RNN model to innovate better performance since the model is able to capture the temporal dependencies and complexities relationships existing between labels. The greater precision of the RNN can be explained by the dynamics of this method as it is able to learn based on previous experiences and make more precise predictions as the next steps continue. Moreover, by definition, RNNs can process sequential data more effectively than their traditional classifiers, and thus they have the benefit of improving the classification accuracy, especially those with many labels of interest, where dependencies in the labels are also important.

Figure 5 compares the precision performance metrics of the new RNN, the current ML-KNN, the FRS-LIFT, the FSEA-MLC, and the ASD-MLC for multi label categorization. The methodologies are shown

on the X-axis in the previous picture, and the precision results are shown on the Y-axis. Based on the results, it can be said that the suggested RNN model, which has a precision of 94.0781%, outperforms the current ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC models, which have respective precisions of 75.4419%, 78.3628%, 82.0435%, and 93.5990% for the Shuttle dataset.

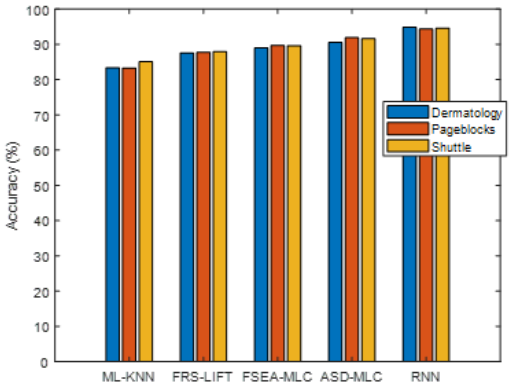

Figure 4. Accuracy results

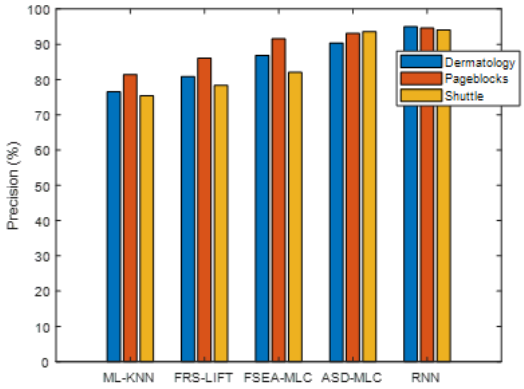

Figure 5. Precision results

Figure 6 displays the comparison outcomesof the existing ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC methods and proposed RNN for multi label classification in terms of recall. The techniques are shown on the X-axis in the previous picture, and the recall results are shown on the Y-axis. The results demonstrated higher recall values of 93.5671% for RNNs when compared to ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC models for the Shuttle dataset, which generate 85.8645%, 87.5996%, 89.5523%, and 92.1842%, respectively. Figure 7 compares the suggested RNN for multi label classification to the current ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC algorithms based on F-measure performance metrics where procedures form the X-axis while their f-measure results from Y-axis. From the outcome, the recommended RNN framework generates greater f-measure results (93.5123%), whereas the current ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC models generate, respectively, only 80.3101%, 82.7194%, 85.7645%, and 92.8362% for the Shuttle dataset.

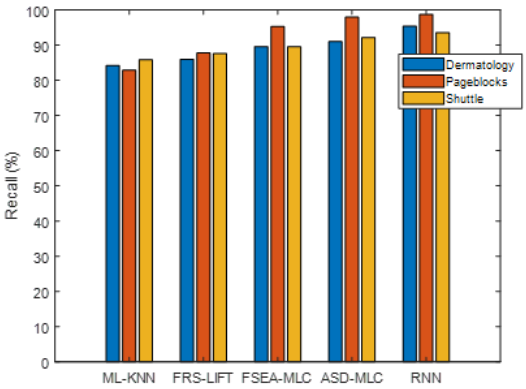

Figure 6. Recall results

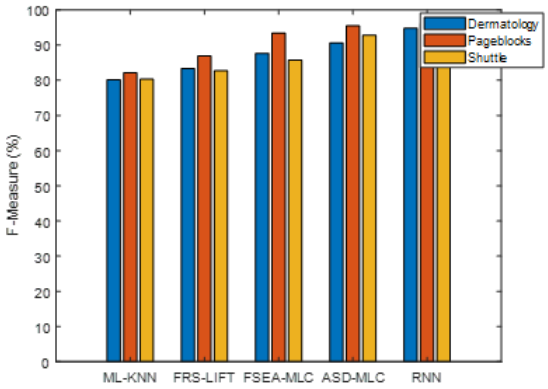

Figure 7. F-measure results

Figure 8 compares the proposed RNN for MLC to the existing ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC algorithms in terms of error rate performance metrics. The illustration up above based on error rate performance metrics where procedures form the X-axis while their error rates form Y-axis. From theoutcomes, the recommended RNN system has a greater error rate of 5.4211%, but the current ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC models only had error rates of 14.9491%, 12.0770%, 10.3799%, and 8.4217%, respectively, for the Shuttle dataset.

Figure 9 shows the specificity performance of the proposed RNN model versus other algorithms like ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC. The RNN model is better specific with high precision of 97.6780% to the Shuttle dataset compared to 81.5660, 83.6670, 94.7890, and 96.7889 of the ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC models respectively.

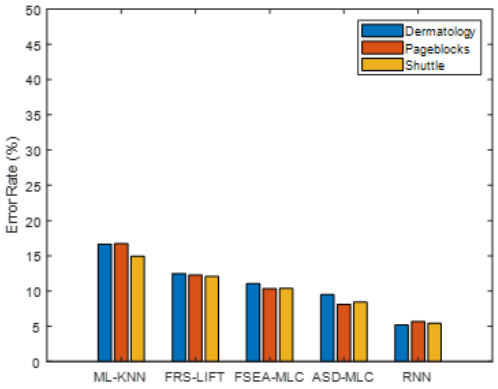

Figure 8. Error rate results

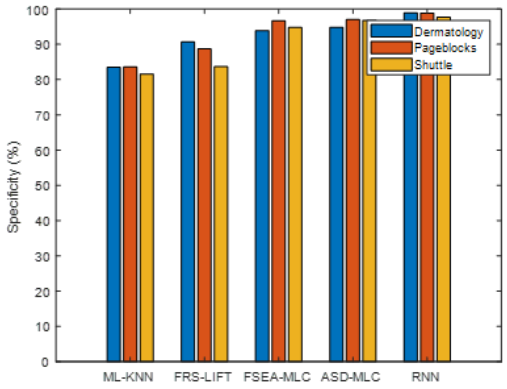

Figure 9. Specificity results

The great specificity of the RNN model can be explained by the capacity to differentiate between true and false negatives and to reduce false positives to minimum. This is particularly so in multi-label classification problems where it is equally important to correctly identify non-relevant labels as it is to predict the relevant labels. Since RNNs have a recurrent structure, they are able to capture better sequential dependencies in sequential data, resulting in improved generalization and reduced miss-classifications. This is because of its capabilities to remember past predictions of labels, which makes the RNN probability to predicate false positives low, therefore enhancing specificity.

5. CONCLUSION

The study outcomes prove that the suggested MLC model with ICSO was effective to select features and the use of RNNs to classify them. The model demonstrated a high accuracy of 94.5789% which is superior over the current classification models, including ML-KNN, FRS-LIFT, FSEA-MLC, and ASD-MLC, in diverse evaluation measures, including precision, recall, F-measure, and specificity.

The use of ICSO to choose features greatly enhanced the performance of the classification process due to the elimination of unnecessary and noisy attributes and the quality of the RNN model relying on the dependency of labels led to its high accuracy and specificity. The results show that ICSO and RNNs can help to deal with the issue of MLC in skewed datasets.

This work not only contributes to the advancement of MLC in imbalanced datasets but also paves the way for its application in critical domains like medical diagnosis and bioinformatics, where accurate and efficient classification is crucial.

Future studies may consider the implementation of this model to other fields, e.g. medical diagnosis and bioinformatics, to provide further support to its use. Also, it would be possible to increase the scalability and ability of the model to bigger and more complex data to further increase its practical applicability.

FUNDING INFORMATION

The authors received no financial support for the research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
M. Priyadharshini	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Narendruni Lakshmi Priya		✓				✓		✓	✓	✓	✓	✓		
Anitha Vippdapu	✓		✓	✓		✓			✓		✓		✓	
Subrata Chowdhury		✓								✓		✓		
Thi-Thu Nguyen		✓								✓		✓		✓
Duc-Tan Tran	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	
Duc-Nghia Tran		✓							✓		✓			✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

We hereby confirm that we have no actual or potential conflicts of interest in regards to this study.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] G. Du, J. Zhang, N. Zhang, H. Wu, P. Wu, and S. Li, "Semi-supervised imbalanced multi-label classification with label propagation," *Pattern Recognition*, vol. 150, p. 110358, 2024, doi: 10.1016/j.patcog.2024.110358.

[2] A. Hashemi, M. B. Dowlatshahi, and H. Nezamabadi-pour, "Minimum redundancy maximum relevance ensemble feature selection: A bi-objective Pareto-based approach," *Journal of Soft Computing and Information Technology*, vol. 12, no. 1, pp. 20–28, 2023.

[3] S. Qin, H. Wu, L. Zhou, Y. Zhao, and L. Zhang, "TAE: Topic-aware encoder for large-scale multi-label text classification," *Applied Intelligence*, vol. 54, no. 8, pp. 6269–6284, 2024, doi: 10.1007/s10489-024-05485-z.

[4] T. Yamaguchi and Y. Yamashita, "Multi-target regression via target combinations using principal component analysis," *Computers & Chemical Engineering*, vol. 181, p. 108510, 2024, doi: 10.1016/j.compchemeng.2023.108510.

[5] J. Y. Hang and M. L. Zhang, "Collaborative learning of label semantics and deep label-specific features for multi-label classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 9860–9871, 2021, doi:10.1109/TPAMI.2021.3136592.

[6] D. Y. Eroglu and U. Akcan, "An adapted ant colony optimization for feature selection," *Applied Artificial Intelligence*, vol. 38, no. 1, p. 2335098, 2024, doi: 10.1080/08839514.2024.2335098.

[7] M. K. Xie and S. J. Huang, "Partial multi-label learning with noisy label identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3676–3687, 2021, doi: 10.1109/TPAMI.2021.3059290.

[8] M. Khodabandeh, A. Mahmoodzadeh, and H. Agahi, "Recognition of PRI modulation using an optimized convolutional neural network with a gray wolf optimization based on internet protocol and optimal extreme learning machine," *Scientific Reports*, vol. 15, no. 1, p. 33760, 2025, doi: 10.1038/s41598-025-89994-y.

[9] Z. Xiong *et al.*, "DRPCA-Net: Make robust PCA great again for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1-16, 2025, doi: 10.1109/TGRS.2025.3588392.

[10] X. Liu, J. He, Z. Wang, and C. Mi, "Smooth robust principal component analysis based on multidimensional transform tensor for dynamic MRI," *Signal Processing*, vol. 227, p. 109712, 2025, doi: 10.1016/j.sigpro.2024.109712.

[11] B. Chen, L. Cao, C. Chen, Y. Chen, and Y. Yue, "A comprehensive survey on the chicken swarm optimization algorithm and its applications: State-of-the-art and research challenges," *Artificial Intelligence Review*, vol. 57, no. 7, p. 170, 2024, doi: 10.1007/s10462-024-10786-3.

[12] L. Yao, J. Yang, P. Yuan, G. Li, Y. Lu, and T. Zhang, "Multi-strategy improved sand cat swarm optimization: global optimization and feature selection," *Biomimetics*, vol. 8, no. 6, p. 492, 2023, doi: 10.3390/biomimetics8060492.

[13] K. Wu, L. Wang, and M. Liu, "ADVCSO: Adaptive Dynamically Enhanced Variant of Chicken Swarm Optimization for Combinatorial Optimization Problems," *Biomimetics*, vol. 10, no. 5, p. 303, 2025, doi: 10.3390/biomimetics10050303.

[14] A. O. Aseeri, "Effective RNN-based forecasting methodology design for improving short-term power load forecasts," *Journal of Computational Science*, vol. 68, p. 101984, 2023, doi: 10.1016/j.jocs.2023.101984.

[15] F. W. Mugware, T. Ravele, and C. Sigauke, "Short-Term Predictions of Global Horizontal Irradiance using advanced ensemble approaches," *Computation*, vol. 13, no. 3, p. 72, 2025, doi: 10.3390/computation13030072.

[16] A. Abedi and S. S. Khan, "FedSL: Federated split learning on distributed sequential data in recurrent neural networks," *Multimedia Tools and Applications*, vol. 83, no. 10, pp. 28891–28911, 2024, doi: 10.1007/s11042-023-15184-5.
[17] V. Veerasamy *et al.*, "Blockchain-based decentralized frequency control of microgrids using federated learning fractional-order recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 1089–1102, 2023, doi: 10.1109/TSG.2023.3267503.
[18] M. Ławryńczuk and K. Zarzycki, "LSTM and GRU type recurrent neural networks in model predictive control: A review," *Neurocomputing*, vol. 632, p. 129712, 2025, doi: 10.1016/j.neucom.2025.129712.
[19] M. Mujahid *et al.*, "An efficient ensemble approach for Alzheimer's disease detection using deep learning," *Diagnostics*, vol. 13, no. 15, p. 2489, 2023, doi: 10.3390/diagnostics13152489.
[20] D. Anelli, P. Morano, F. Tajani, and M. R. Guarini, "The interpretative effects of normalization techniques on complex regression modeling," *Information*, vol. 16, no. 6, p. 486, 2025, doi: 10.3390/info16060486.
[21] KEEL Dataset Repository, Knowledge Extraction based on Evolutionary Learning, 2004-2018, [Online]. Available: https://sci2s.ugr.es/keel/imbalanced.php?order=featRsub10. (Accessed: Mar. 03, 2026).
[22] S. A. Alex, J. J. V. Nayahi, and S. Kaddoura, "Deep convolutional neural networks with genetic algorithm-based SMOTE for imbalanced data classification," *Applied Soft Computing*, vol. 156, p. 111491, 2024, doi: 10.1016/j.asoc.2024.111491.
[23] Q. WRuan and W. X. WHuang, "Feature selection via label enhancement and neighborhood rough set for multi-label data with unbalanced distribution," *Applied Soft Computing*, vol 175, 2025, doi: 10.1016/j.asoc.2025.113028.
[24] H. Chen *et al.*, "Compound strategy based binary willow catkin optimization for feature selection," *Cluster Computing*, vol. 28, no. 3, 2025, doi: 10.1007/s10586-024-04879-5.
[25] A. Jasmir, P. A. Jusia, Y. Arvita, and G. Gunardi, "Comparative analysis of word embedding features to improve the performance of deep learning models on social media data," *Bulletin of Electrical Engineering and Informatics*, vol 14, no 4, pp. 2923-2934, 2025, doi: 10.11591/eei.v14i4.9200.

BIOGRAPHIES OF AUTHORS

M. Priyadharshini received her Ph.D. degree in Computer Science from Bharathiar University, Coimbatore, India in 2022. She is working as an Associate Professor in the Department of Computer Science and Engineering at ICFAI Foundation for Higher Education, Hyderabad, India. She has 14 years of academic experience. She has 9 years of research experience. She has published books, book chapters and, research articles in reputed international journals. She has organized national and international-level events. She has been a resource person for guest lectures, seminars, and faculty development programs. She also reviewed and evaluated more than 20 papers from conferences and journals, book chapters, and science articles in data mining, machine learning, AI, data science, IoT, and blockchain. She has attended many FDPs, workshops, and webinars. She has published many papers, copyrights, and patents in her name. She can be contacted at email: mpriyadharshinimca@gmail.com.

Narendruni Lakshmi Priya is working as degree lecturer in MJPTBCWRDC(W), Station Ghanpur, Suryapet, Hyderabad, India. She has 16 years of academic experience. She has 8 years of research experience. She has published research articles in reputed international journals. She has organized national and international-level events. She has been a resource person for guest lectures, seminars, and faculty development programs. She has attended many FDPs, workshops, and webinars. She has published many papers, copyrights, and patents in her name. She can be contacted at email: nlp.priya.66@gmail.com.

Anitha Vippdapu is working as an Assistant Professor in the Department of Computer Science and Engineering at ICFAI Foundation for Higher Education, Hyderabad, India. She has 16 years of Teaching experience. She has 6 years of research experience. She has published research articles in reputed international journals. She has attended many FDPs, workshops, and webinars. She has published many papers, copyrights, and patents in her name. She can be contacted at email: anitha.vippadapu@gmail.com.

Subrata Chowdhury is an Associate Professor in the Department of Master of Computing Application at C. Abdul Hakeem College of Engineering and Technology, Tamil Nadu, India. He received his MCA from VIT Vellore, M.Tech. from JNTU Anantapur, and Ph.D. in Computer Science from VISTAS, Chennai. He is currently pursuing a Ph.D. at NIT Arunachal Pradesh. His research interests include artificial intelligence, machine learning, IoT, and cloud computing. He has published over 125 research papers, authored and edited several books, holds an H-index of 22, and serves as a reviewer for Springer, Elsevier, Taylor and Francis, and MDPI. He can be contacted at email: subrata895@gmail.com.

Thi-Thu Nguyen received the Engineer degree in electrical Engineering from University of Transport and Communications, Vietnam, in 2000. She received the Master degree in Telecommunications Engineering from University of Engineering and Technology, Vietnam in 2005 and Ph.D degree in Electronic Engineering from Le Quy Don Technical University, Vietnam in 2018. Currently she is lecturer in Ha Noi University of Industry. Her research interests include the areas of deep learning, artificial intelligence, space-time signal processing for communications such as MIMO, spatial modulation, and cooperative communications. She can be contacted at email: thunt@haui.edu.vn.

Duc-Tan Tran received B.Tech. degree in electronics engineering from College of Technology, Hanoi, Vietnam in 2002 and M.Tech. degree in electronics engineering from College of Technology, Hanoi, Vietnam in 2005 and Ph.D. degree in electronics engineering from electronics engineering VNU University of Engineering and Technology, Hanoi, Vietnam. He is currently Vice Dean of Faculty of Electrical and Electronics Engineering, PHENIKAA University. His current research interest is in signal processing for biomedical imaging. He can be contacted at email: tan.tranduc@phenikaa-uni.edu.vn.

Duc-Nghia Tran received the Ph.D. degree from Sorbonne Paris Cité, France. In his thesis, he focuses on signal processing of EPR spectra for in vivo experiments. He is a researcher with the Institute of Information Technology (IoIT), Vietnam Academy of Science and Technology (VAST). He has published about 20 research articles. His publication received the Best Paper Award from the International Conference on Advanced Technologies for Communications (ATC'17). His research interests include mathematics and signal processing, electron paramagnetic resonance (EPR), parameter estimation, and data analysis. He serves as a Technical Committee Program Member and a reviewer for many international conferences and journals. He can be contacted at email: nghiatd@ioit.ac.vn.