

Automatic Model for Postpartum Depression Identification using Deep Reinforcement Learning and Differential Evolution Algorithm

Sunyuan Shen^{1*}, Sheng Qi², Hongfei Luo³

Department of Applied Engineering, Zhejiang Business College Hangzhou 310053, People's Republic of China^{1,3}
Institute of Electronic Commerce, Zhejiang Business College Hangzhou, 310053, People's Republic of China²

Abstract—Postpartum depression (PPD) affects approximately 12% of new mothers, posing a significant health concern for both the mother and child. However, many women with PPD do not receive proper care. Preventative interventions are more cost-effective for high-risk women, but identifying those at risk can be challenging. To address this problem, we present an automatic model for PPD using a deep reinforcement learning approach and a differential evolution (DE) algorithm for weight initialization. DE is known for its ability to search for global optima in high-dimensional spaces, making it a promising approach for weight initialization. The policy of the model is based on an artificial neural network (ANN), treating the categorization issue as a policymaking stage-by-stage process. The DE algorithm is used to acquire initial weight values, with the agent obtaining samples and performing classifications in each step. The habitat provides an award for every categorization activity, considering a greater award for identification of the minor category to encourage precise detection. By using a particular compensatory technique and an encouraging learning system, the operator eventually decides the most excellent method for achieving its goals. The model's efficiency is evaluated by analyzing a set of data acquired from the population-based BASIC study carried out in Uppsala, Sweden, which covers the period from 2009 to 2018 and consists of 4313 samples. The experiential results, identified by known analysis criteria, indicate that the sample achieved better precision and correctness, making it suitable for identifying PPD. The proposed model could have significant implications for identifying at-risk women and providing timely interventions to improve maternal and child health outcomes.

Keywords—Postpartum depression; deep reinforcement learning; differential evolution algorithm; weight initialization; artificial neural network

I. INTRODUCTION

PPD is a common condition in Sweden, affecting 8% to 15% of new mothers annually [1]. It manifests as mild to severe depressive episodes either during pregnancy or within the first year after giving birth [2, 3]. The exact cause of PPD remains unknown but is thought to result from a combination of psychosocial, psychological, and biological factors. Biologically, inflammation, the withdrawal of allopregnanolone, and genetic factors play roles. Psychosocially, factors like ongoing stress, prior depression, relationship difficulties, and significant life changes contribute to PPD risk. The consequences of PPD can be severe,

impacting both mother and child. Mothers may struggle with forming emotional bonds with their child, doubt their caregiving abilities, and even have harmful thoughts towards the child [4]. Efforts have been made to predict PPD during the prenatal period. Still, currently, there is no reliable method to accurately identify women at risk of experiencing depressive symptoms after giving birth [5].

Conventional statistical methods typically analyze the relationship between two variables while factoring in other variables [6, 7]. In contrast, machine learning (ML) techniques allow for the simultaneous analysis of many interconnected variable relationships, leading to the creation of data-driven predictive models [8]. These models can then be assessed to find the most effective predictor. ML can handle complex nonlinear relationships and integrate various data types from different sources. Over the past ten years, the application of ML has expanded across medical fields including oncology, cardiology, hematology, critical care, and psychiatry. In PPD, which poses a moderate risk of a serious psychiatric condition with reasonably accurate prediction of symptom onset, ML can be highly valuable given the societal impact of PPD. Despite its potential benefits, it is impractical to monitor every individual for early PPD symptoms. A more efficient strategy is to target high-risk groups during postpartum checks by healthcare professionals like midwives or nurses, rather than the broader population. In Sweden, with its 120,000 annual births and the myriad of post-childbirth adjustments women undergo, and a typical PPD prevalence of around 12%, this targeted approach proves especially advantageous for personalized, cost-effective maternal and perinatal mental care.

Machine learning can face issues with feature extraction, affecting generalization, processing time, and precision [9]. The rise of deep learning, particularly Multi-Layer Perceptron (MLP), offers improved classification capabilities [10]. MLP, designed for nonlinear XOR problems, is versatile for various tasks, from image processing to optimization [11]. It functions like human neurons, where each node in ANN processes inputs and uses an activation function to produce an output. In MLP, nodes are interconnected across different layers, without intra-layer connections.

Medical classification poses significant challenges due to imbalanced data, where negative instances far outnumber positive ones, leading to decreased performance [9, 12, 13].

Measures can be employed at both the algorithm and data levels to address this issue. At the level of the data, downsampling, upsampling, or a mixture of both techniques can be utilized to alleviate the negative impact of imbalanced classification [14, 15]. On the other hand, algorithmic approaches involve assigning greater weight to the minority class [16, 17]. Moreover, deep learning methods offer potential solutions for tackling imbalanced classification [12, 18]. Huang et al. [19] have proposed a process to identify distinctive features in imbalanced data while maintaining inter-cluster and inter-class margins. Similarly, Yan et al. [16] have suggested a technique using the bootstrapping method to balance data in convolutional networks across mini-batches.

Population-based training can be utilized to select the most optimal solution from a population of generated models in order to optimize neural networks [20-22]. This approach mitigates the risk of being trapped in local optima, a common challenge in traditional training methods [23]. Surprisingly, a straightforward evolutionary algorithm has proven comparable to stochastic gradient descent in terms of the effectiveness of neural network training [24, 25]. Jaderberg et al. [26] successfully applied population-based training to cutting-edge models in deep reinforcement learning, machine translation, and generative adversarial networks, yielding consistent accuracy, training time, and stability enhancements. In related studies [27] and [28], effective weight training for neural networks was achieved through the adoption of differential evolution-based strategies [29] and the employment of the ABC (Artificial Bee Colony) method [30, 31], respectively.

This paper introduces a novel approach for identifying PPD by combining deep Q-learning and the DE method to initialize the load. The categorization task is formulated as an estimating challenge within an RL framework, treating it as a Markov decision process. The environment state is represented by a sample, and the agent is an ANN. To initiate the game, we explore the application of the DE algorithm to find an optimal weight initialization for the ANN. The agent classifies each sample, and its classification is awarded accordingly by giving the right choices positive awards and wrong decisions getting negative awards. For tackling dataset imbalance, the minority class is given a higher absolute value of the reward. The operator aims to amplify the accumulative awards by accurately classifying the samples throughout the policymaking procedure. Significantly, our research is pioneering in utilizing a population-based methodology that leverages an extensive and varied dataset, incorporating a wide range of clinical, psychometric self-report, and medical journal-derived variables. The performance of our proposed model on this dataset demonstrates its supremacy over alternative methods depending on initializing the arbitrary load. The primary contributions of the paper can be outlined in the following manner:

- Formulating the classification task as a guessing game within an RL framework, treating it as a Markov decision process.
- Using DE to find an optimal weight initialization for the ANN, initiating the guessing game.

- Rewarding correct and incorrect decisions positively, addressing dataset imbalance by giving higher rewards to the minority class.
- Demonstrating the superiority of the proposed model over alternative approaches that rely on random weight initialization through its performance on the dataset.

The organization of this paper is outlined below: Section II reviews relevant literature, Section III delves into the DE algorithm, and Section IV describes the proposed model. Results and their analysis are discussed in Section V. The paper concludes with a summary in Section VI, along with recommendations for future investigations.

II. RELATED WORK

In recent years, the field of medical science has witnessed an unprecedented surge in the application of machine learning techniques to forecast and categorize a plethora of health concerns, with PPD standing out as a significant area of interest [32]. To understand the evolution and progression of these methodologies, a series of pioneering studies have been meticulously evaluated to shed light on the practices adopted and the degree of precision achieved in their PPD classification endeavors [33].

Zhang et al. [34] placed their bet on SVM and FFS-RF, emphasizing these as the most promising tools for PPD prediction. They embarked on a comprehensive longitudinal survey, engaging 508 women as respondents. The Edinburgh Postnatal Depression Scale (EPDS) served as their choice of instrument to gauge PPD risk. Delving further into their work, Zhang et al. [35] opted for EHR datasets, focusing on the detection of PPD in perinatal women. Their findings were intriguing; while logistic regression fortified with L2 regularization emerged as the top contender for data leading up to childbirth, the post-childbirth data saw MLP taking the lead. Jasiya et al. [36] proposed a machine learning system to identify risk factors and prevalence of postpartum depression in Bangladesh. Utilizing modified questions from EPDS and PHQ-2 scales and socio-demographic queries, data from 150 women was analyzed. The most effective model was identified as Random Forest. Amit et al. [37] presented a Gradient Boosting Machine (GBM)-based approach to PPD depression risk using electronic health records from 266,544 UK women between 2000 and 2017. The model assessed socio-demographic and medical variables and was evaluated alongside the standard EPDS questionnaire for improved screening accuracy. Park et al. [38] suggested an evaluation of methods to reduce bias in clinical machine learning models. Health data from the IBM MarketScan Medicaid Database, focusing on females aged 12 to 55 years with a live birth record from 2014 to 2018, was analyzed. The study examined logistic regression, random forest, and extreme gradient boosting models for postpartum depression and mental health service utilization, assessing racial disparities. Bias reduction methods like reweighing and Prejudice Remover were also explored.

Diversifying the landscape, Shin et al. [39] ventured to harness the PRAMS 2012-2013 dataset and the PHQ-2 questionnaire. Their objective was clear: to tap into various

machine learning algorithms and decipher the prevalence of PPD. Their rigorous analysis crowned Random Forest as the algorithm par excellence for PPD prediction. In a similar vein, Andersson et al. [40] crafted multiple machine learning prototypes, drawing data from Swedish hospitals. Their study, vast in its scope, found the Extremely Randomized Trees model to be unmatched in performance.

Dipping into another significant contribution, Tortajada et al. [41] embarked on a study, leveraging data accumulated from hospital settings. Their research, hinging on MLP, showcased an impressive accuracy rate of 81% in PPD prediction. On the other hand, De Choudhury [42] ventured into the realm of digital platforms, conducting a longitudinal online survey. This study examined an array of regression models, paving the way for new insights. In another noteworthy study, Natarajan et al. [43] showcased a comparative analysis of algorithms like Functional-gradient boosting, Decision-trees, Naive Bayes, and SVM. Their work, rooted in a longitudinally curated dataset, championed Functional-gradient boosting as the superior method. Lastly, Wang et al. [44] married EHR data with machine learning techniques, with their research revealing SVM as the most fitting algorithm for their dataset.

However, while these advancements are commendable, it is essential to recognize the challenges that come with them. The sheer diversity of algorithms means that selecting the optimal one requires rigorous testing, often demanding substantial resources. Moreover, discrepancies in datasets across different studies might lead to varying conclusions, underscoring the need for standardized and universally accepted data collection methods. Additionally, the robustness of these models in real-world scenarios remains a topic of debate, necessitating further in-depth research and validation.

III. DIFFERENTIAL EVOLUTION

Differential Evolution (DE) [29] is an optimization algorithm that stems from populations and finds frequent applications in addressing optimization problems. It falls under the umbrella of evolutionary algorithms, drawing inspiration from the natural progression of evolution. DE is widely acknowledged for its straightforwardness and effectiveness in handling optimization problems involving continuous variables. In addition to its broad range of applications, DE has proven valuable in the realm of machine learning, particularly within the domain of training artificial neural networks. An essential aspect of neural network training revolves around weight initialization, a pivotal factor influencing convergence, generalization capabilities, and the capacity to learn intricate patterns. Traditional weight initialization methods, such as random initialization or fixed values, often grapple with the challenge of striking an optimal balance between avoiding vanishing or exploding gradients and achieving efficient learning. DE can be employed to initialize neural network weights by treating weight values as variables to be optimized. The objective is to identify an optimal set of weight values that minimize the objective function, representing network performance or error on a training dataset. Through the strategic reimagining of weight initialization as an optimization quandary, DE can adeptly

navigate and investigate weight configurations that serve as a catalyst for bolstering network performance. DE offers several benefits for weight initialization in machine learning [10]:

- **Exploration of Solution Space:** The DE algorithm facilitates the exploration of the solution space by generating diverse candidate solutions. This is particularly advantageous for weight initialization as it helps to avoid getting stuck in local optima and enables the algorithm to search for better-performing weight configurations.
- **Efficient Optimization:** The DE algorithm optimizes the weights by iteratively updating them based on the difference between the target and current solutions. This efficient optimization process aids in finding suitable initial weights that can contribute to faster convergence and improved learning algorithm performance.
- **Robustness to Noise:** The DE algorithm is known for its robustness to noisy fitness evaluations. In weight initialization, this robustness helps to handle uncertainties and variations in the data, leading to more reliable and stable initial weight configurations.
- **Flexibility and Adaptability:** The DE algorithm allows for flexibility and adaptability in weight initialization. It can be customized to handle specific problem domains or constraints, such as imposing bounds on weight values or incorporating prior knowledge. This adaptability enhances the algorithm's ability to initialize weights suitable for the given learning task.

The primary procedures of DE are as follows:

- **Initialization:** The method begins with creating a beginning populace of chosen resolutions called "individuals." Each individual represents a potential solution to the optimization problem and is usually represented as a vector of real numbers.
- **Mutation:** In each iteration of the algorithm, the individuals in the population are subjected to mutation. Mutation is the process of generating new candidate solutions by perturbing existing ones. In DE, the mutation is performed by creating a trial vector for each individual using the difference between two randomly selected individuals from the population.
- **Crossover:** After mutation, a crossover operation is applied to combine the trial vector with the original individual. Crossover is a process that blends the information from the trial vector and the original individual to create a new candidate solution. The crossover operation in DE is typically performed using a binomial crossover scheme, where each component of the new solution is selected either from the trial vector or the original individual with a certain probability.
- **Selection:** The new candidate solution produced by crossover is compared with the original individual, and the better one is selected to proceed to the next iteration. The selection process ensures that only the fitter individuals survive and propagate their traits to the next

generation. This step helps in driving the search towards better solutions over time.

- Termination: The algorithm continues to iterate through mutation, crossover, and selection until a termination condition is met. The termination condition can be a maximum number of iterations, reaching a desired level of solution quality, or any other criteria defined by the problem.

IV. MODEL ARCHITECTURE

To address our research challenge, we turned to the capabilities of DE for weight initialization and RL for imbalanced classification, particularly because existing models fall short in several aspects. Traditional models often rely on random weight initialization, which can lead to prolonged training times and the possibility of converging to suboptimal solutions. Many existing algorithms struggle with imbalanced datasets, leading to biased predictions that often overlook the minority class.

By using DE for weight initialization, we can ensure a diverse and potentially more optimal starting point for our learning algorithm. This can lead to faster convergence and potentially superior solutions compared to traditional methods.

RL is a type of machine learning wherein an agent learns to decide by taking actions in an environment to maximize a cumulative reward. It has especially apt for imbalanced classification tasks because it can be tailored to place greater emphasis on the minority class by suitably adjusting the reward mechanism. In scenarios where traditional supervised learning faces challenges because of insufficient representative data for all classes, RL can more effectively explore the decision space and devise strategies that prioritize the accurate classification of underrepresented classes. This

addresses another critical limitation of many existing models: their inability to adapt to and accurately classify instances from underrepresented categories.

A. Pretraining

Weight initialization is crucial in neural network training, influencing convergence, generalization, and pattern learning. In this article, the DE algorithm treats weights as variables and minimizes the objective function, representing performance or error. DE effectively explores the weight space by iteratively evaluating and updating weight configurations through evolutionary operators. The goal is to refine weights for better convergence, reduced error, and improved generalization. Incorporating DE in pretraining enhances weight initialization, improving overall network performance.

In this article, the power of the DE algorithm is leveraged to initiate the weights of the MLP. The weights are encoded by meticulously arranging them into a vector, representing them within the DE algorithm. It should be noted that finding the most appropriate layout can be intricate, requiring persistent efforts and a multitude of experiments. Undeterred by the complexity, an optimal encoding strategy was devised through extensive trials and refinements. To provide a visual depiction of this process, Fig. 1 is presented, offering a clear illustration of how all the weights and bias terms are meticulously gathered and assembled into a comprehensive vector. This vector, acting as a candidate solution within the DE algorithm, encapsulates the essential components necessary for weight initialization. By organizing these elements thoughtfully and strategically, the stage is set for the DE algorithm to unleash its optimization prowess, guiding the MLP toward enhanced performance and increased learning capability.

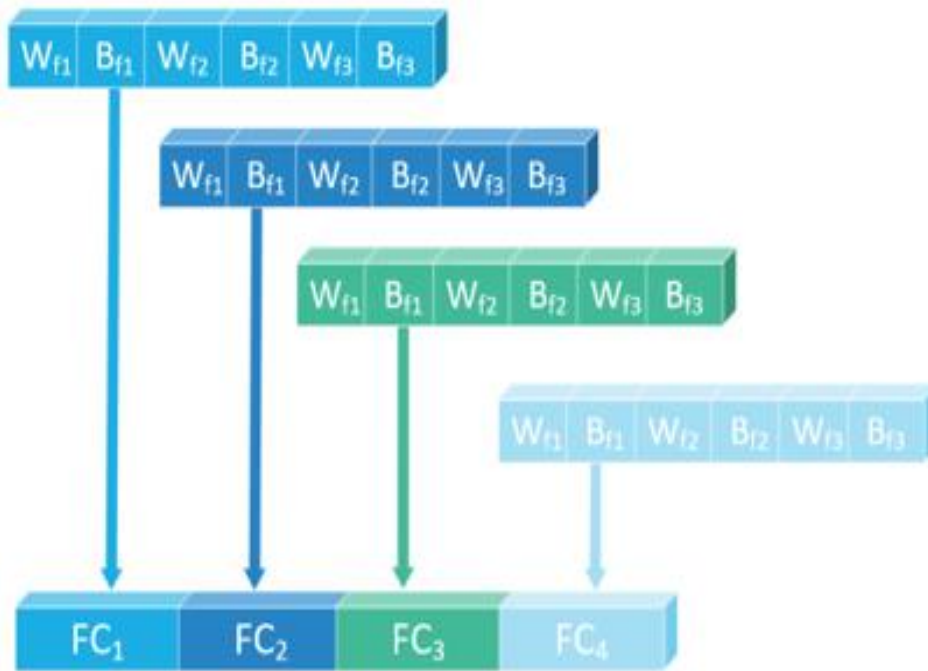


Fig. 1. Encoding strategy used in the proposed algorithm.

The efficacy of a chosen resolution is comprehensively assessed by describing and establishing athleticism performance as a crucial metric. Within the context of the specific problem domain, the quality and performance of the solution are quantified through this function, which serves as a vital tool. The fitness function is meticulously crafted to capture the essential aspects and criteria that govern the success and effectiveness of the candidate solution. The fitness or suitability of the solution can be objectively measured through the careful consideration of relevant factors and parameters, allowing informed decisions to be made and the optimization process to be guided towards optimal outcomes. The fitness function is defined as:

$$Fitness = \frac{1}{\sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (1)$$

Here, N shows the whole count of train demos, where y_i represents the goal value of the i -th sample and \hat{y}_i shows the corresponding output predicted by the model.

B. Prediction

To further improve our method of calling the problem of unbalanced classification caused by unequal data volumes in our two classes, we implemented a consecutive policymaking procedure using an RL method. This involved training an ANN model to act as an agent, making informed classifications for each instance, and effectively handling the challenges associated with imbalanced datasets. In the sequential decision-making process, each instance in the train dataset represented a distinct habitat state. The ANN model, acting as the operator, made a categorization sequence for every instance. Simultaneously, as the operator predicted the category name, for instance, it took an act denoted as a_t . At each time-step t , the agent observed an instance representing the current state of the environment, labeled as s_t . The environment provided a reward, r_t , in response to the agent's actions, aiming to guide its behavior. To address the class imbalance issue, the reward values were carefully crafted. Samples from the majority class received lower absolute reward values, while relatively higher absolute reward values were assigned to samples from the minority class. This reward design aimed to encourage the agent to prioritize the correct classification of minority class samples, contributing to mitigating the impact of imbalanced data. In this article, the reward function is defined as:

$$r_t(s_t, a_t, y_t) = \begin{cases} +1, & a_t = y_t \text{ and } s_t \in D_S \\ -1, & a_t \neq y_t \text{ and } s_t \in D_S \\ \lambda, & a_t = y_t \text{ and } s_t \in D_H \\ -\lambda, & a_t \neq y_t \text{ and } s_t \in D_H \end{cases} \quad (2)$$

where D_S , and D_H represent the minority and majority categories in order. Incorrectly/correctly categorizing a demo of the major category gains an award of $+\lambda/-\lambda$, where $0 < \lambda < 1$. We aimed to incentivize the agent to give greater attention to the minority class and mitigate the bias caused by imbalanced data by providing differential rewards based on class distribution. Through this reinforcement learning approach, the agent learned an optimal classification strategy that considered both the inherent difficulty of

classifying the minority class and the importance of accurate predictions overall. During training, the agent continuously refined its decision-making capabilities and updated its policies and strategies based on the rewards received. By leveraging reinforcement learning techniques, we aimed to achieve a more balanced and effective classification performance, particularly for the underrepresented class. This novel sequential decision-making process enabled us to overcome the limitations imposed by imbalanced datasets and successfully address the challenges of imbalanced classification. As a result, we achieved improved accuracy and fairness in predictions by combining the power of artificial neural networks and reinforcement learning.

V. EXPERIMENTAL RESULTS

A. Data Sources

The data used for developing the prediction models were acquired from the "Biology, Affect, Stress, Imaging and Cognition during Pregnancy and the Puerperium" (BASIC) study [45]. BASIC is a prospective cohort study conducted at the Department of Obstetrics and Gynaecology in Uppsala University Hospital, Uppsala, Sweden, and it involves a population-based approach. Between September 2009 and November 2018, pregnant women who fulfilled specific eligibility criteria were invited to take part in the study. The criteria included being 18 years of age or older, not having concealed identities, possessing sufficient proficiency in reading and comprehending Swedish, and not having been diagnosed with bloodborne infections or non-viable pregnancies based on routine ultrasound examinations. In the BASIC study, data collection primarily relied on online surveys and questionnaires administered to women at various stages: during pregnancy at the 17th and 32nd week of gestation, as well as at 6 weeks, 6 months, and 12 months after giving birth. These surveys and questionnaires were designed to gather information from participants during these specific time points. The surveys consisted of inquiries regarding various background characteristics, encompassing sociodemographic variables, psychological assessments, medical details, reproductive history, lifestyle factors, and sleep patterns. All questionnaires were completed by the participants themselves and were conducted online. Information was additionally sourced from medical journals. The study had a participation rate of 20%, but the cohort experienced a comparatively low dropout rate, as 71% of the participants remained in the study during the 12-month follow-up period. The study obtained approval from the Research Ethics Board in Uppsala (Dnr 2009/171, with amendments). Prior to their inclusion in the study, all participating women provided written informed consent. The research methods adhered to applicable guidelines and regulations.

B. Model Performance

Our project utilizes a 64-bit Windows operating system, complemented with 64 GB of RAM and a 64 GB GPU. Table I presents the hyperparameters applied to the proposed model.

TABLE I. HYPERPARAMETER SETTING FOR THE PROPOSED MODEL

Hyperparameter	Value
Epoch	256
Batch size	64
Learning rate	0.02
Dropout rate	0.5
Discount factor	0.3

TABLE II. HYPERPARAMETER SETTING FOR MACHINE LEARNING METHODS

Algorithm	Parameter	Value
Naïve Bayes	α (Lidstone smoothing parameter)	0.5
KNN	k (Number of neighbors)	5
	Distance Metric	Euclidean ($p=2$), Manhattan ($p=1$)
SVM	Kernel	polynomial
	γ (Kernel coefficient)	0.5
Random Forests	Number of trees	20
	Max depth of tree	10
Logistic Regression	C (Inverse regularization strength)	0.3
	Solver	liblinear
Decision Tree	Criterion	entropy
	Max depth	10

In the evaluation process, the suggested model underwent rigorous comparison with six distinct machine learning models: Naïve Bayes [46], KNN [47], SVM [48], Random forests [49], Logistic Regression [50], and Decision tree [51]. In the evaluation process, the suggested model underwent rigorous comparison with six distinct machine learning models: Naïve Bayes [46], KNN [47], SVM [48], Random forests [49], Logistic Regression [50], and Decision tree [51]. Table II shows the parameters applied to these models.

Additionally, two modified versions were included in the analysis to explore different variations of the proposed model. The first modified version, proposed+random weights, adopted a similar foundational architecture to our model but employed random weights for initialization. This alternative initialization method allowed for a comparative investigation of the impact of weight initialization on the performance of the model. The second modified version, Proposed+random weights+RL, incorporated RL techniques for classification. This integration of RL aimed to enhance the ability of the model to make accurate predictions and improve its overall performance. Standard metrics were employed to assess these models' performance, with particular emphasis on the

geometry average and F-measure due to their suitability for unbalanced info [52]. The results, which can be found in Table III, clearly demonstrate the superiority of the proposed model over all other models, including the previously recognized top performer, Decision Tree. The evaluation results are shown schematically in Fig. 2 to understand the results better. Across all evaluation criteria, the proposed model consistently outperformed its counterparts. Notably, the proposed model achieved remarkable error reductions over 65%; plus, 29% in the G-averages and F-measure metrics, orderly. These substantial improvements illustrate the effectiveness of the proposed model in tackling the challenges posed by imbalanced data and its ability to generate more accurate predictions. Contrasting the offered sample with the modified versions, Offered+arbitrary loads+RL and Offered+arbitrary loads, the significance of the integration of DE and RL approaches becomes apparent. Our model demonstrated an impressive decrease in the error rate of approximately 62% when compared to these modified versions. This finding underscores the critical role played by DE and RL in enhancing the model's performance and highlights their importance in developing state-of-the-art machine learning models.

TABLE III. OUTCOMES OF SEVERAL CATEGORIZATION METHODS

	accuracy	recall	precision	F-measure	G-means
Naïve Bayes	0.6804±0.0501	0.5909±0.0653	0.5353±0.1206	0.5601±0.0652	0.6562±0.0052
KNN	0.8101±0.1553	0.7351±0.0043	0.7653±0.1002	0.7471±0.1054	0.8002±0.0402
SVM	0.7803±0.1005	0.6603±0.0054	0.6903±0.0452	0.6702±0.2103	0.7503±0.0051
Random forests	0.6901±0.1006	0.5602±0.2202	0.5453±0.1001	0.5502±0.0953	0.6505±0.2502
Logistic Regression	0.8105±0.1404	0.7706±0.0456	0.7005±0.1003	0.7302±0.1404	0.8006±0.1201
Decision tree	0.8304±0.1405	0.8103±0.1006	0.7502±0.2704	0.7902±0.2001	0.8202±0.0104
Proposed+random weights	0.8104 ± 0.0627	0.8202 ± 0.1108	0.8013 ± 0.0109	0.7918 ± 0.1623	0.8303 ± 0.2622
Proposed+random weights+RL	0.8615 ± 0.0243	0.8704 ± 0.1256	0.8509 ± 0.2691	0.8506 ± 0.0517	0.8609 ± 0.0921
Proposed	0.8907 ± 0.0384	0.9053 ± 0.1132	0.8847 ± 0.0315	0.8844 ± 0.0297	0.90423

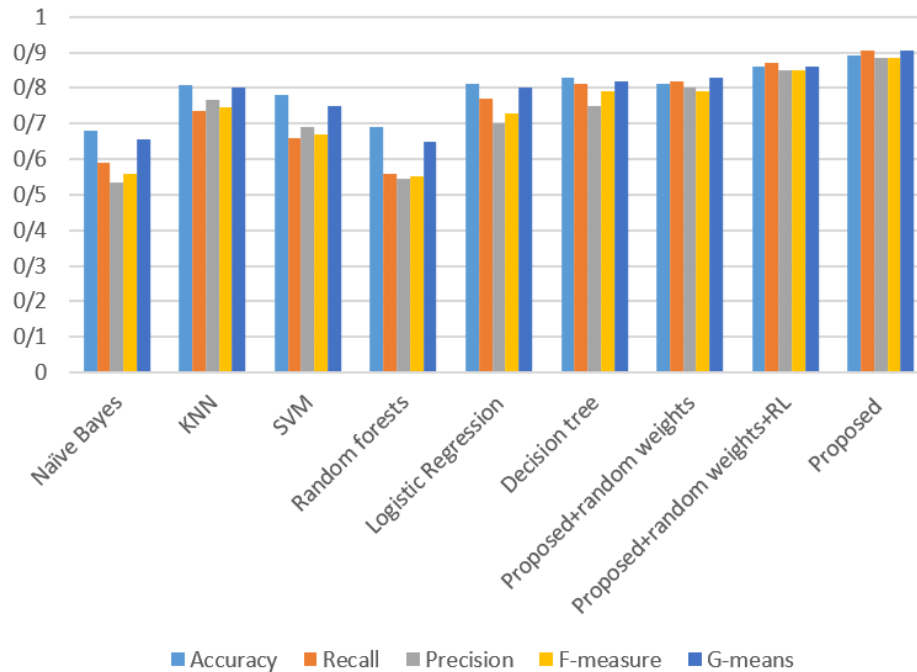


Fig. 2. Graphical comparison of various classification algorithms.

TABLE IV. HYPERPARAMETER SETTING FOR METAHEURISTIC ALGORITHMS

Algorithm	Parameter	Value
DE	scaling factor	0.5
	crossover probability	0.7
ABC	limit	$n_e \times \text{dimensionality}$
	n_o	50% of the colony
	n_e	50% of the colony
	n_s	1
FA	light absorption coefficient	1
	attractiveness at $r = 0$	0.2
	scaling factor	0.25
BA	constant for loudness update	0.5
	constant for an emission rate update	0.5
	initial pulse emission rate	0.001
COA	discovery rate of alien solutions	0.25

TABLE V. OUTCOMES OF SEVERAL METAHEURISTIC METHODS

	accuracy	recall	precision	F-measure	G-means
Proposed + ABC + RL	0.8510 ± 0.1470	0.8390 ± 0.1220	0.8570 ± 0.0500	0.8400 ± 0.0080	0.8190 ± 0.4110
Proposed + GWO+ RL	0.8400 ± 0.1560	0.8290 ± 0.1010	0.8400 ± 0.2490	0.8280 ± 0.0140	0.7980 ± 0.0230
Proposed + FA+ RL	0.8250 ± 0.0020	0.8130 ± 0.1210	0.8090 ± 0.2600	0.8170 ± 0.0630	0.7780 ± 0.0000
Proposed + BA+ RL	0.8080 ± 0.0120	0.7980 ± 0.0040	0.8000 ± 0.0590	0.8030 ± 0.1430	0.7600 ± 0.1180
Proposed + COA+ RL	0.7900 ± 0.1570	0.7700 ± 0.0120	0.7840 ± 0.2590	0.7700 ± 0.1640	0.7390 ± 0.1480

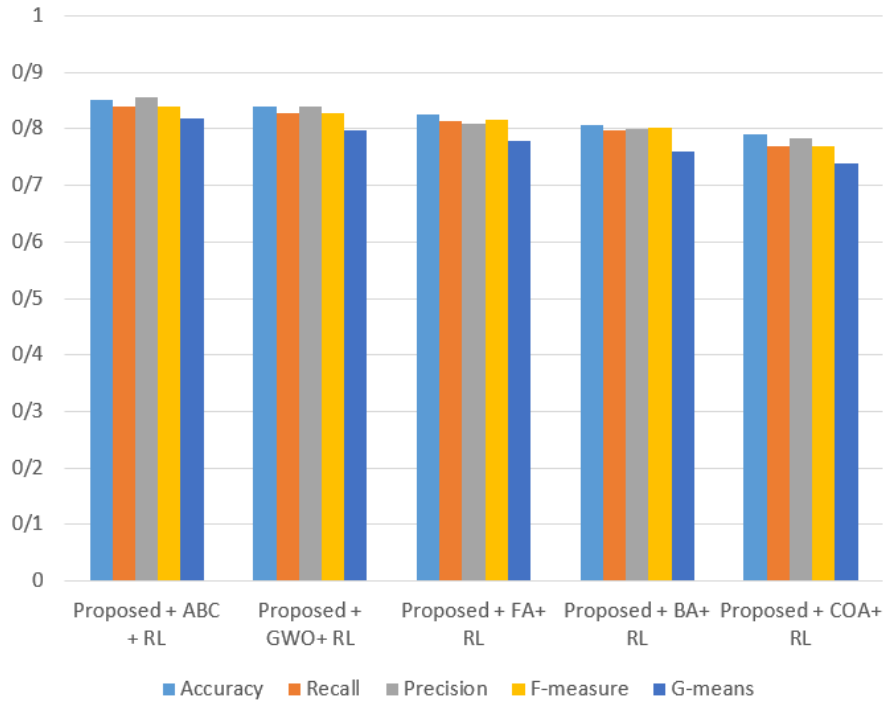


Fig. 3. Graphical comparison of various classification algorithms.

A detailed analysis was conducted in the subsequent experiment to compare the DE algorithm with various well-established metaheuristic optimization algorithms. To ensure a fair comparison, different metaheuristics were employed to derive the initial weights while keeping the remaining components of the model consistent. The evaluation encompassed six distinct algorithms, namely ABC [53], GWO [54], FA [55], BA [56], and COA [57]. For every algorithm, both the population size and the count of function evaluations are configured to 200 and 3,000, respectively. The default configurations are detailed in Table IV. The results of this comprehensive experiment were systematically presented in Table V and Fig. 3, providing valuable insights into the performance of each algorithm. Notably, the findings highlighted the remarkable achievement of DE, which demonstrated a significant reduction in error of approximately 52% when compared to the ABC algorithm. Furthermore, the DE algorithm outperformed other well-known algorithms, including GWO and BA. This outcome solidified the position of the DE algorithm as a leading contender among the considered metaheuristic optimization approaches.

C. Award Operation Effect

The rewards given to the majority and minority classes for correct and incorrect classifications are +1 and $\pm\lambda$, in order. The λ value is determined by the scale of major to minor demos, and by increasing this scale, the ideal value of λ is expected to decrease. For researching the λ effect, we assessed the offered demo's efficiency through various λ on a scale of 0 to 1 (in increments of 0.1) as saving the bonus for the major category. The outcomes are represented in Fig. 4. By the time the λ is 0, the major category's effect gets insignificant, while at $\lambda = 1$, both categories have the same effects. The findings reveal the model performs optimally when λ is set to 0.4 for every measured criterion, recommending that the optimum λ value is on the scale of 0 to 1. We should notice that when it is crucial to diminish the major category's effect by adjusting λ , adjusting it to a lower level might have a detrimental effect on the overall model performance. The results indicate that the choice of λ has a substantial impact on the performance of the model. The optimal value of λ depends on the relative proportions of the majority and minority samples, underscoring the significance of careful selection to achieve the best possible outcomes.

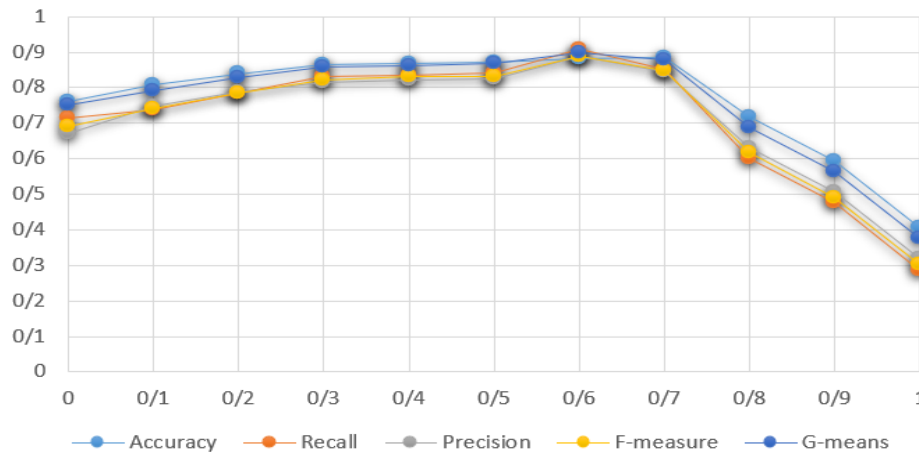


Fig. 4. Visual depiction showcasing the alteration in performance parameters caused by fluctuations in the value of λ .

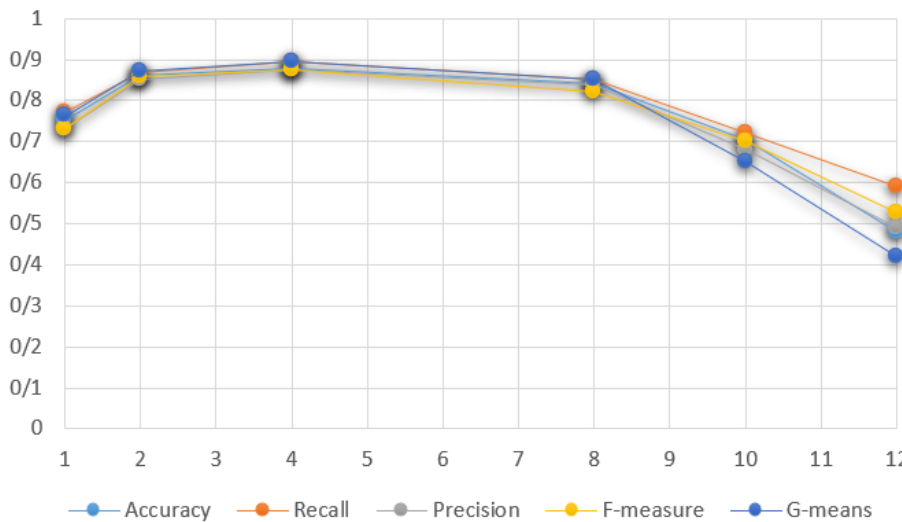


Fig. 5. The plotted performance metrics as a function of the MLP layers.

D. Impact of the MLP Layers

The article emphasizes that increasing the number of layers in an MLP leads to a higher model complexity, which in turn increases the risk of overfitting. On the other hand, having too few layers may limit the ability of the model to capture important features in the training data. In our proposed approach, we conducted experiments with six different values (1, 2, 4, 8, 10, 12) for the number of layers in the MLP to examine its impact on model performance. The results, presented in Fig. 5, demonstrate a decreasing trend in performance as the number of layers is in the 1 to 4 range, next to a rising trend for 4 to 12 values. This suggests that having four layers in the MLP yields optimal performance and achieves the best results.

E. Impact of the Loss Function

Various techniques are available to tackle data imbalances in machine learning models, including adjusting data augmentation methods and selecting an appropriate loss function. Among these techniques, the choice of loss function plays a crucial role in enabling the model to learn from the minority class effectively. To assess the efficacy of different

loss functions, we examined five specific functions: WCE [58], BCE [59], DL [60], TL [61], and CL [62]. BCE and WCE commonly use loss functions that equally treat positive and negative examples. However, in the case of imbalanced datasets where the emphasis needs to be placed on the minority class, these loss functions may not be suitable.

On the other hand, DL and TL loss functions are better suited for imbalanced datasets as they yield improved performance on the minority class. As a promising loss function, CL is particularly beneficial for applications involving unbalanced data. By adjusting the weights of the loss function, CL can assign lower importance to simple examples and focus more on learning complex samples. To evaluate the effectiveness of these loss functions, we conducted experiments and presented the results in Table 6 and Fig. 6. The findings demonstrate that the CL function surpasses the TL function, resulting in a 25% reduction in the error rate for the accuracy metric and a 39% reduction for the F-measure metric. However, it is worth noting that the CL function performs 60% worse than the FL function, which is a specialized loss function specifically designed for binary classification tasks. It is important to consider these results in

the context of the specific problem at hand and the nature of the dataset. While the CL function outperforms the TL function, it falls short when compared to the FL function. Further investigation is required to understand the factors contributing to these differences in performance and to explore the potential of customized loss functions specifically

tailored to address the challenges posed by imbalanced datasets. Additionally, research can focus on developing novel loss functions or adapting existing ones to strike a balance between emphasizing the minority class and maintaining overall classification accuracy across various classification tasks and datasets.

TABLE VI. RESULTS FROM VARIOUS LFS

	accuracy	recall	precision	F-measure	G-means
WCE	0.7432± 0.1132	0.7200± 0.0100	0.7174± 0.0155	0.7215± 0.0130	0.7510± 0.1255
BCE	0.8001± 0.1021	0.7830± 0.1145	0.7520± 0.1051	0.7622± 0.0102	0.8025± 0.0101
DL	0.8042± 0.2033	0.8151± 0.0025	0.7936± 0.0106	0.7992± 0.0106	0.826± 0.1000
TL	0.8320± 0.1203	0.8252± 0.0006	0.8014± 0.0216	0.8148± 0.0436	0.8450± 0.2062
CL	0.8630± 0.0247	0.8545± 0.2045	0.8415± 0.0148	0.8440± 0.0152	0.8639± 0.0152

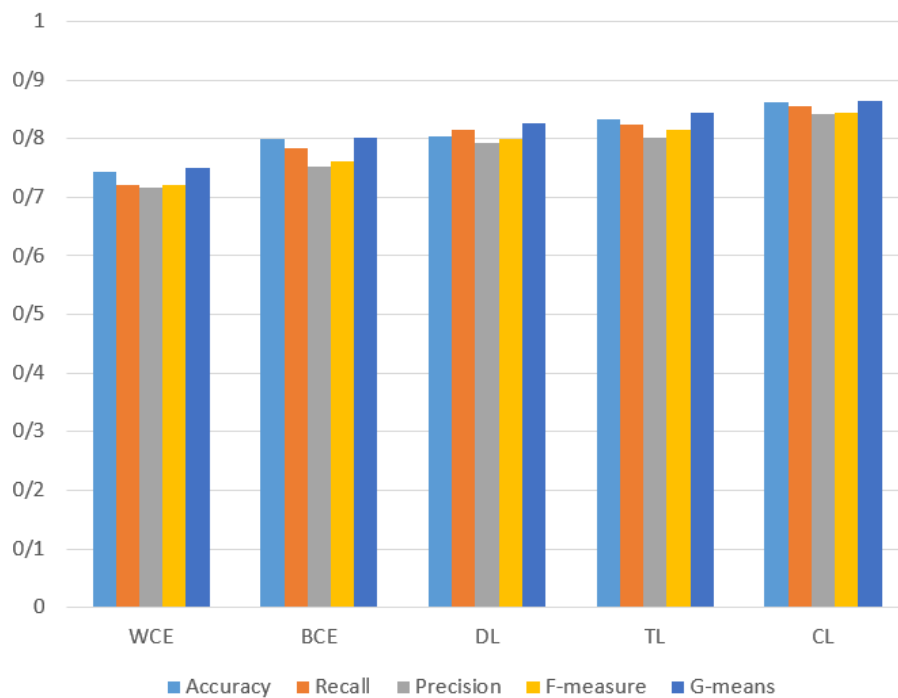


Fig. 6. Graphical comparison of various loss functions.

F. Discussion

The findings presented in this study have important implications for PPD identification and intervention. Developing an automated model using a deep reinforcement learning approach and a DE algorithm for weight initialization shows the potency of advanced ML approaches for addressing the challenges associated with PPD identification. One of the key advantages of using the DE algorithm for weight initialization is its ability to explore high-dimensional spaces and find optimal weight values effectively. This ensures that the model is initialized in a manner that enables it to make accurate predictions and classify PPD effectively. By leveraging an ANN and treating the categorization issue as a policymaking stage-by-stage process, the sample considers the complexity and nuances of PPD, enhancing its predictive capabilities. Using an encouraging learning system and a

particular compensatory technique further enhances the model's performance. By assigning a higher reward for identifying the minority class, the model is incentivized to focus on precise detection, addressing the challenge of identifying at-risk individuals. This approach acknowledges the importance of early identification to provide timely interventions and support to those most in need. The evaluation of the model's performance using a comprehensive dataset acquired from the population-based BASIC study in Uppsala, Sweden, strengthens the validity of the findings. With a large sample size of 4313 samples spanning a significant period, the study provides robust evidence of the high accuracy of the model in identifying PPD. This accuracy underscores the potential effectiveness of the model in real-world applications for identifying at-risk women. The implications of this research are far-reaching. By accurately

identifying women at risk for PPD, healthcare professionals can provide timely interventions and support, thus improving maternal and child health outcomes. Preventative interventions targeted at high-risk individuals have been shown to be more cost-effective, making the automated model a valuable tool in resource allocation and optimizing healthcare services.

To offer a more in-depth assessment of our model's capabilities, we reached out for external expert opinions. By teaming up with seasoned professionals, possessing extensive experience in the field, we embarked on a comprehensive qualitative review of the model's performance. These experts, hailing from diverse backgrounds and having a rich tapestry of experiences in similar research areas, thoroughly scrutinized the model's underpinnings, methodologies, and outcomes. Their rigorous evaluations and constructive feedback painted a clear picture. Their collective insights resoundingly echoed our preliminary findings, particularly highlighting the model's unparalleled precision, steadfast reliability, and robust adaptability. When our model was placed side by side with pre-existing algorithms for a comparative analysis, it distinctly stood out, showcasing its superior design and performance. The external validation from such esteemed professionals not only fortified our confidence in the model but also underscored its potential for real-world applications and future research endeavors.

However, it is important to acknowledge the limitations of this study and consider avenues for future research. Firstly, the dataset used in this study was acquired from a specific population-based study conducted in Uppsala, Sweden. While this provides valuable insights into the model's performance within that particular context, it raises questions about the generalizability of the findings to diverse populations and settings [63]. Variations in cultural, socioeconomic, and healthcare factors may influence the prevalence and presentation of PPD, potentially impacting the model's performance [64]. Therefore, future studies should aim to validate the model using datasets from different regions and populations to ensure its applicability across various contexts [65].

Additionally, while the model demonstrates high accuracy in identifying PPD, assessing its performance in real-world clinical arrangements is crucial [66]. The controlled environment of the study may not fully reflect the complexities and challenges faced by healthcare professionals in their daily practice.

Evaluating the model's effectiveness in a clinical setting, where multiple factors can influence the identification and treatment of PPD, would provide valuable insights into its practical utility. Longitudinal studies tracking patient outcomes and the impact of the model's predictions on treatment decisions and health outcomes would further enhance our understanding of its clinical relevance. Furthermore, expanding the scope of research beyond the model accuracy is essential. While accuracy is a crucial metric, evaluating other performance measures such as sensitivity, specificity, positive predictive value, and negative predictive value is equally important [67]. These metrics

provide a more comprehensive assessment of the model's diagnostic capabilities and ability to identify individuals at risk and those not at risk for PPD. Understanding the model performance across these measures can guide healthcare professionals in effectively utilizing its predictions and making informed interventions and resource allocation decisions.

Moreover, assessing the impact of the model on patient outcomes is a critical aspect that requires further investigation [68]. While timely identification of at-risk women is essential, evaluating whether the interventions based on the model predictions lead to improved maternal and child health outcomes is equally vital [69]. Conducting studies that measure the effectiveness of interventions guided by the model, such as targeted support programs or personalized treatment plans, would provide valuable evidence of the model's potential to impact patient outcomes positively.

Finally, it is important to consider the ethical implications and potential challenges associated with the implementation of an automated model for PPD identification [70]. Issues such as privacy, data security, and the potential for biases in the model's predictions need to be thoroughly examined and addressed. Ensuring transparency, fairness, and accountability in developing and deploying such models is essential to maintaining trust among healthcare professionals and the wider public [71].

VI. CONCLUSION

In this study, we have developed an automated model to identify PPD using a deep reinforcement learning approach combined with a DE algorithm for weight initialization. The DE algorithm is renowned for its ability to effectively explore high-dimensional spaces and find optimal weight values, making it well-suited for weight initialization in our model. Our approach utilizes an ANN and treats the PPD classification problem as a policymaking stage-by-stage process. At every stage, the operator acquires samples and employs classifications, while the habitat maintains rewards for every classifying activity. For inspiring precise detection, a greater award is determined for recognizing the minor category. Through a particular compensatory technique and an encouraging learning system, the operator learns and chooses the most effective method for achieving the goals. To evaluate our sample's efficiency, we analyzed a comprehensive set of data obtained from the population-based BASIC study conducted in Uppsala, Sweden, spanning from 2009 to 2018, and comprising 4313 samples. The experiential results were assessed by known analysis criteria, revealing our sample achieved greater precision and correctness, demonstrating its suitability for identifying PPD. These findings carry significant implications for identifying at-risk women and providing timely interventions to improve maternal and child health outcomes.

Reinforcement learning algorithms often face the challenge of striking a balance between exploration and exploitation. Future research can delve deeper into exploring effective strategies for addressing this trade-off in the context of PPD identification. Techniques such as adaptive exploration policies, multi-objective optimization, or

incorporating domain knowledge can help optimize the model's performance in identifying at-risk women while minimizing false positives and negatives. Moreover, investigating the potential of transfer learning and domain adaptation techniques can contribute to improving the generalization capabilities of the PPD identification model. By leveraging knowledge gained from related domains or pre-trained samples, the sample's efficiency can be enhanced when applied to different populations, cultures, or healthcare settings. This research direction can help address the challenges of model generalizability and make the automated PPD identification model more robust.

REFERENCES

- [1] L. J. Miller, "Postpartum depression," *Jama*, vol. 287, no. 6, pp. 762-765, 2002.
- [2] I. S. Yim, L. R. Tanner Stapleton, C. M. Guardino, J. Hahn-Holbrook, and C. Dunkel Schetter, "Biological and psychosocial predictors of postpartum depression: systematic review and call for integration," *Annual review of clinical psychology*, vol. 11, pp. 99-137, 2015.
- [3] M. Bloch, R. C. Daly, and D. R. Rubinow, "Endocrine factors in the etiology of postpartum depression," *Comprehensive psychiatry*, vol. 44, no. 3, pp. 234-246, 2003.
- [4] D. K. Sit and K. L. Wisner, "The identification of postpartum depression," *Clinical obstetrics and gynecology*, vol. 52, no. 3, p. 456, 2009.
- [5] S. Asif et al., "Severe obstetric lacerations associated with postpartum depression among women with low resilience—a Swedish birth cohort study," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 127, no. 11, pp. 1382-1390, 2020.
- [6] E. Y. Wan, C. A. Moyer, S. D. Harlow, Z. Fan, Y. Jie, and H. Yang, "Postpartum depression and traditional postpartum care in China: role of zuoyuezi," *International Journal of Gynecology & Obstetrics*, vol. 104, no. 3, pp. 209-213, 2009.
- [7] P. Tong, L.-P. Dong, Y. Yang, Y.-H. Shi, T. Sun, and P. Bo, "Traditional Chinese acupuncture and postpartum depression: a systematic review and meta-analysis," *Journal of the Chinese Medical Association*, vol. 82, no. 9, pp. 719-726, 2019.
- [8] Y. Lee et al., "Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review," *Journal of affective disorders*, vol. 241, pp. 519-532, 2018.
- [9] S. V. Moravvej et al., "RLMD-PA: A reinforcement learning-based myocarditis diagnosis combined with a population-based algorithm for pretraining weights," *Contrast Media & Molecular Imaging*, vol. 2022, 2022.
- [10] S. V. Moravvej, S. J. Mousavirad, D. Oliva, G. Schaefer, and Z. Sobhaninia, "An improved de algorithm to optimise the learning process of a bert-based plagiarism detection model," in *2022 IEEE Congress on Evolutionary Computation (CEC)*, 2022: IEEE, pp. 1-7.
- [11] S. Zhang, C. Tjortjis, X. Zeng, H. Qiao, I. Buchan, and J. Keane, "Comparing Data Mining Methods with Logistic Regression."
- [12] S. Danaei et al., "Myocarditis Diagnosis: A Method using Mutual Learning-Based ABC and Reinforcement Learning," in *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo)*, 2022: IEEE, pp. 000265-000270.
- [13] H. Zareiamand, A. Darroudi, I. Mohammadi, S. V. Moravvej, S. Danaei, and R. Alizadehsani, "Cardiac Magnetic Resonance Imaging (CMRI) Applications in Patients with Chest Pain in the Emergency Department: A Narrative Review," *Diagnostics*, vol. 13, no. 16, p. 2667, 2023.
- [14] I. Mani and I. Zhang, "kNN approach to unbalanced data distributions: a case study involving information extraction," in *Proceedings of workshop on learning from imbalanced datasets*, 2003, vol. 126: ICML, pp. 1-7.
- [15] S. V. Moravvej, M. J. M. Kahaki, M. S. Sartakhti, and A. Mirzaei, "A method based on attention mechanism using bidirectional long-short term memory (BLSTM) for question answering," in *2021 29th Iranian Conference on Electrical Engineering (ICEE)*, 2021: IEEE, pp. 460-464.
- [16] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in *Advances in Intelligent Computing: International Conference on Intelligent Computing, ICIC 2005*, Hefei, China, August 23-26, 2005, *Proceedings, Part I 1*, 2005: Springer, pp. 878-887.
- [17] L. Hong et al., "GAN-LSTM-3D: An efficient method for lung tumour 3D reconstruction enhanced by attention-based LSTM," *CAAI Transactions on Intelligence Technology*, 2023.
- [18] M. Soleimani, Z. Forouzanfar, M. Soltani, and M. J. Harandi, "Imbalanced Multiclass Medical Data Classification based on Learning Automata and Neural Network," *EAI Endorsed Transactions on AI and Robotics*, vol. 2, 2023.
- [19] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5375-5384.
- [20] S. Moravvej, M. Maleki Kahaki, M. Salimi Sartakhti, and M. Joodaki, "Efficient GAN-based method for extractive summarization," *Journal of Electrical and Computer Engineering Innovations (JECEI)*, vol. 10, no. 2, pp. 287-298, 2022.
- [21] S. V. Moravvej, S. J. Mousavirad, M. H. Moghadam, and M. Saadatmand, "An LSTM-based plagiarism detection via attention mechanism and a population-based approach for pre-training parameters with imbalanced classes," in *Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part III 28*, 2021: Springer, pp. 690-701.
- [22] M. Bahadori, M. Soltani, M. Soleimani, and M. Bahadori, "Statistical Modeling in Healthcare: Shaping the Future of Medical Research and Healthcare Delivery," in *AI and IoT-Based Technologies for Precision Medicine: IGI Global*, 2023, pp. 431-446.
- [23] M. Soltani, M. Bahadori, and M. Soleimani, "Optimal Predictive Maintenance and Spare Part Inventory Policies for a Degrading System Subjected to Imperfect Actions," Available at SSRN 4558570.
- [24] S. V. Moravvej, M. Joodaki, M. J. M. Kahaki, and M. S. Sartakhti, "A method based on an attention mechanism to measure the similarity of two sentences," in *2021 7th International Conference on Web Research (ICWR)*, 2021: IEEE, pp. 238-242.
- [25] S. V. Moravvej, A. Mirzaei, and M. Safayani, "Biomedical text summarization using conditional generative adversarial network (CGAN)," *arXiv preprint arXiv:2110.11870*, 2021.
- [26] M. Jaderberg et al., "Population based training of neural networks," *arXiv preprint arXiv:1711.09846*, 2017.
- [27] S. J. Mousavirad, D. Oliva, S. Hinojosa, and G. Schaefer, "Differential evolution-based neural network training incorporating a centroid-based strategy and dynamic opposition-based learning," in *2021 IEEE Congress on Evolutionary Computation (CEC)*, 2021: IEEE, pp. 1233-1240.
- [28] D. Karaboga, B. Akay, and C. Ozturk, "Artificial bee colony (ABC) optimization algorithm for training feed-forward neural networks," in *Modeling Decisions for Artificial Intelligence: 4th International Conference, MDAI 2007, Kitakyushu, Japan, August 16-18, 2007. Proceedings 4*, 2007: Springer, pp. 318-329.
- [29] S. V. Moravvej, S. J. Mousavirad, D. Oliva, and F. Mohammadi, "A Novel Plagiarism Detection Approach Combining BERT-based Word Embedding, Attention-based LSTMs and an Improved Differential Evolution Algorithm," *arXiv preprint arXiv:2305.02374*, 2023.
- [30] S. Vakilian, S. V. Moravvej, and A. Fanian, "Using the artificial bee colony (ABC) algorithm in collaboration with the fog nodes in the Internet of Things three-layer architecture," in *2021 29th Iranian Conference on Electrical Engineering (ICEE)*, 2021: IEEE, pp. 509-513.
- [31] M. S. Sartakhti, M. J. M. Kahaki, S. V. Moravvej, M. javadi Joortani, and A. Bagheri, "Persian language model based on BiLSTM model on COVID-19 corpus," in *2021 5th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, 2021: IEEE, pp. 1-5.

- [32] S. R. Low, S. A. Bono, and Z. Azmi, "The effect of emotional support on postpartum depression among postpartum mothers in Asia: A systematic review," *Asia-Pacific Psychiatry*, p. e12528, 2023.
- [33] K. M. Deligiannidis et al., "Zuranolone for the treatment of postpartum depression," *American Journal of Psychiatry*, vol. 180, no. 9, pp. 668-675, 2023.
- [34] W. Zhang, H. Liu, V. M. B. Silenzio, P. Qiu, and W. Gong, "Machine learning models for the prediction of postpartum depression: application and comparison based on a cohort study," *JMIR medical informatics*, vol. 8, no. 4, p. e15516, 2020.
- [35] Y. Zhang, S. Wang, A. Hermann, R. Joly, and J. Pathak, "Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women," *Journal of affective disorders*, vol. 279, pp. 1-8, 2021.
- [36] J. F. Raisa, M. S. Kaiser, and M. Mahmud, "A machine learning approach for early detection of postpartum depression in Bangladesh," in *International Conference on Brain Informatics, 2022*: Springer, pp. 241-252.
- [37] G. Amit et al., "Estimation of postpartum depression risk from electronic health records using machine learning," *BMC Pregnancy and Childbirth*, vol. 21, no. 1, pp. 1-10, 2021.
- [38] Y. Park et al., "Comparison of methods to reduce bias from clinical prediction models of postpartum depression," *JAMA network open*, vol. 4, no. 4, pp. e213909-e213909, 2021.
- [39] D. Shin, K. J. Lee, T. Adeluwa, and J. Hur, "Machine learning-based predictive modeling of postpartum depression," *Journal of Clinical Medicine*, vol. 9, no. 9, p. 2899, 2020.
- [40] S. Andersson, D. R. Bathula, S. I. Iliadis, M. Walter, and A. Skalkidou, "Predicting women with depressive symptoms postpartum with machine learning methods," *Scientific reports*, vol. 11, no. 1, p. 7877, 2021.
- [41] S. Tortajada et al., "Prediction of postpartum depression using multilayer perceptrons and pruning," *Methods of information in medicine*, vol. 48, no. 03, pp. 291-298, 2009.
- [42] !!! INVALID CITATION !!! {}.
- [43] S. Natarajan, A. Prabhakar, N. Ramanan, A. Bagilone, K. Siek, and K. Connelly, "Boosting for postpartum depression prediction," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017*: IEEE, pp. 232-240.
- [44] K. L. Wisner, E. L. Moses-Kolko, and D. K. Sit, "Postpartum depression: a disorder in search of a definition," *Archives of women's mental health*, vol. 13, pp. 37-40, 2010.
- [45] S. Andersson, D. R. Bathula, S. I. Iliadis, M. Walter, and A. Skalkidou, "Predicting women with depressive symptoms postpartum with machine learning methods," *Scientific reports*, vol. 11, no. 1, pp. 1-15, 2021.
- [46] G. I. Webb, E. Keogh, and R. Miikkulainen, "Naïve Bayes," *Encyclopedia of machine learning*, vol. 15, pp. 713-714, 2010.
- [47] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003*. Proceedings, 2003: Springer, pp. 986-996.
- [48] M. A. de Almeida, "DATA MINING: DETERMINAC AO DE AGRUPAMENTOS EM GRANDES BASES DE DADOS," *Universidade Federal do Rio de Janeiro*, 2013.
- [49] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5-32, 2001.
- [50] M. P. LaValley, "Logistic regression," *Circulation*, vol. 117, no. 18, pp. 2395-2399, 2008.
- [51] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 18, no. 6, pp. 275-285, 2004.
- [52] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced data sets," *J Inf Eng Appl*, vol. 3, no. 10, 2013.
- [53] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Applied soft computing*, vol. 8, no. 1, pp. 687-697, 2008.
- [54] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in engineering software*, vol. 69, pp. 46-61, 2014.
- [55] X.-S. Yang, "Firefly algorithm, stochastic test functions and design optimisation," *International journal of bio-inspired computation*, vol. 2, no. 2, pp. 78-84, 2010.
- [56] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," *Nature inspired cooperative strategies for optimization (NICSO 2010)*, pp. 65-74, 2010.
- [57] S. Vakilian, S. V. Moravvej, and A. Fanian, "Using the cuckoo algorithm to optimizing the response time and energy consumption cost of fog nodes by considering collaboration in the fog layer," in *2021 5th International Conference on Internet of Things and Applications (IoT), 2021*: IEEE, pp. 1-5.
- [58] Ö. Özdemir and E. B. Sönmez, "Weighted cross-entropy for unbalanced data with application on covid x-ray images," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 2020*: IEEE, pp. 1-6.
- [59] F. Huang, J. Li, and X. Zhu, "Balanced Symmetric Cross Entropy for Large Scale Imbalanced and Noisy Data," *arXiv preprint arXiv:2007.01618*, 2020.
- [60] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, "Dice loss for data-imbalanced NLP tasks," *arXiv preprint arXiv:1911.02855*, 2019.
- [61] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in *Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8, 2017*: Springer, pp. 379-387.
- [62] S. A. Taghanaki et al., "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 24-33, 2019.
- [63] G. J. Domek, L. Heller Szafran, A. Jimenez-Zambrano, and L. Silveira, "Impact on maternal postpartum depressive symptoms of a primary care intervention promoting early language: a pilot study," *Maternal and Child Health Journal*, vol. 27, no. 2, pp. 346-355, 2023.
- [64] V. A. Yakupova, A. D. Suarez, and L. A. Shraibman, "Socioeconomic Risk Factors for Postpartum Depression and Postpartum Post-Traumatic Stress Disorder," *CLINICAL PSYCHOLOGY*, vol. 20, no. 1, pp. 182-201, 2023.
- [65] K. J. Tanna and K. M. Unadkat, "Socio-economic and Psychological Correlates of Postpartum Depression at Six Months," *J Indian Med Assoc*, vol. 121, no. 3, pp. 21-4, 2023.
- [66] K. M. Tabb et al., "Racial differences in immediate postpartum depression and suicidal ideation among women in a Midwestern delivery hospital," *Journal of affective disorders reports*, vol. 1, p. 100008, 2020.
- [67] J. Meijer et al., "Predictive accuracy of Edinburgh postnatal depression scale assessment during pregnancy for the risk of developing postpartum depressive symptoms: a prospective cohort study," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 121, no. 13, pp. 1604-1610, 2014.
- [68] B. D. Thombs et al., "Depression screening and patient outcomes in pregnancy or postpartum: a systematic review," *Journal of psychosomatic research*, vol. 76, no. 6, pp. 433-446, 2014.
- [69] M. E. Gerbasi et al., "Associations between commonly used patient-reported outcome tools in postpartum depression clinical practice and the Hamilton Rating Scale for Depression," *Archives of Women's Mental Health*, vol. 23, pp. 727-735, 2020.
- [70] M. Conway and D. O'Connor, "Social media, big data, and mental health: current advances and ethical implications," *Current opinion in psychology*, vol. 9, pp. 77-82, 2016.
- [71] L. Hidalgo-Padilla, M. Toyama, J. H. Zafrá-Tanaka, A. Vives, and F. Diez-Canseco, "Association between maternity leave policies and postpartum depression: a systematic review," *Archives of Women's Mental Health*, vol. 26, no. 5, pp. 571-580, 2023.