

# Optimization of Feature Weighting for Epitope Classification in B-Cell and SARS Using TVIWACRI-PSO-ELM

1<sup>st</sup> Imam Cholissodin

Department of Informatics Engineering  
Universitas Brawijaya  
Malang, Indonesia  
imamcs@ub.ac.id

2<sup>nd</sup> Nanik Suciati

Department of Informatics  
Institut Teknologi Sepuluh Nopember  
Surabaya, Indonesia  
nanik@if.its.ac.id

3<sup>rd</sup> Darlis Herumurti

Department of Informatics  
Institut Teknologi Sepuluh Nopember  
Surabaya, Indonesia  
darlis@if.its.ac.id

4<sup>th</sup> Riyanarto Sarno

Department of Informatics  
Institut Teknologi Sepuluh Nopember  
Surabaya, Indonesia  
riyanarto@if.its.ac.id

5<sup>th</sup> Valentina Yurina

Department Of Pharmacy  
Universitas Brawijaya  
Malang, Indonesia  
v\_yurina@ub.ac.id

**Abstract**— In the bio-molecular field, epitope classification is essential in vaccine development. A machine learning-based approach has been used for epitope classification using peptide data from b-cell, Severe Acute Respiratory Syndrome (SARS) to predict material for SARS-CoV-2 vaccine. This study aims to improve the performance of epitope classification by weighting peptide features and including input weight and biases of Classifier model using the Time Variant Inertia Weight, Acceleration Coefficients and Random Injection-Particle Swarm Optimization (TVIWACRI-PSO) and the Extreme Learning Machine (ELM) classification algorithm. Experiment on b-cell, SARS dataset shows that feature weighting (FW) using TVIWACRI-PSO can improve accuracy performance by 8.4%.

**Keywords**—*epitope b-cell and sars, sars-cov-2 vaccine, tviwacri-psy-elm, feature weight*

## I. INTRODUCTION

In critical and endemic conditions of Covid-19 and its new variants as of today, studies in determining the prediction of the best peptide epitopes as vaccine ingredients are still necessitated for development [1]–[6], especially the ones grown through recombinant bacteria, both originating only from the spike protein or a combination, for example with a trivalent vaccine design [7]–[10]. This technique is very susceptible to failure in its manufacture if the selected epitope is not correct or is still easily mutated. This is also based on its relation to health since the recovery from diseases caused by the SARS-CoV-2 virus and the risk on human life when treatment is carried out through vaccination activities with a fairly high risk [11]–[15]. Among the difficulties in determining the epitopes of peptides, the most basic and important is that it is very difficult to determine the correct weight value for each data feature manually. As a result, it is definitely not enough to carry out an initial analysis or a standard process by selecting the features using only a subset of the selected features [16]–[19], or by calculating and knowing the correlation value of each feature to the target class or just making a visualization of the data distribution of each feature to the target [20]–[25]. The proper step urgently required is to be able to automate assigning the appropriate weights to each feature as part of the fair treatment process. Moreover, it is also related to the classification algorithm used which is actually good and fast, for example using standard

ELM to multi-layer based [26], [27]. However, it is susceptible to random input weight values, is biased and is sometimes trapped in the results of singular matrix operations [28]–[30].

In 2019, optimization through the incorporation of PSO-ELM for the stabilization of base materials through the process of predicting material performance was able to provide structural quality and durability for flexibility and stiffness of materials in a flexible manner when compared to the conventional method, with higher evaluation results when compared to the combination of PSO and Artificial Neural Network (ANN) and KELM [31], method. However, it still uses TVIW and TVAC, which have a constant value, for each iteration, which tends to cause an imbalance in the exploration and exploitation process [32]–[34]. Then in the research on medical data that tends to have high dimensions with computational complexity constraints, PSO is used for dimension reduction in the form of feature selection which is proven to have the ability to increase efficiency and effectiveness in the form of increasing accuracy values [35]. In the development of PSO in recent studies, however, there is a tendency to use feature weighting in continuous value intervals [0;1] so that feature selection with only binary values is also included in it. In 2021, PSO is utilized to estimate the search for model parameter values as a prediction of the number of death cases caused by COVID-19 based on the logistic function, the solution of which is difficult because it involves quadratic programming [36]. This hyperparameter is very synonymous with optimizing the use of data such as hyperdata which can use methods such as SMOTE [37]. Nevertheless, the SMOTE technique is limited to overly manipulating data. In 2023, there is research that discusses the prediction of drought severity level in the state of Tamil Nadu using 62 years of unbalanced data as a successful effort for geographical mitigation activities using 2 main parameters, i.e., standardized precipitation index (SPI) and standardized precipitation evapotranspiration index (SPEI). PSO is used for feature weighting before being included in the Random Forest (RF) classification method which has succeeded in increasing the precision value by 45% [38].

The contribution of the research presented focuses on hyperparameters on features rather than index data, which are

based on several related and relevant previous studies. The reason is that the entire dataset used is still sufficient to be directly computed with existing computing resources. The focus of the first contribution is on optimizing the feature weights of peptide data from b-cell, SARS, and sars-cov-2 by using the Parallel Time Variant (PTVPSO) algorithm with dynamic TVIW and TVAC in each regeneration which provides balanced conditions in local and global searches and is given an additional random injection (RI) scheme so that it is not easily trapped in early convergence and local optimum. The second contribution is provided to the optimization of input weights and bias in the ELM classification algorithm. The optimization of both is to overcome trial constraints which are generally carried out manually or only a few times when performed without optimization techniques.

## II. METHODS

### A. Features weighting and tuning hyperparameter

Each learning process on the algorithm to obtain the best model results is generally like the one shown in the illustration of Fig. 1. At the beginning of the process, it starts with pre-processing which can use a certain approach according to the dataset specifications [39]–[44], then it is followed by Feature weighting (FW) and hyperparameter tuning (HP). Feature weighting (FW) is a technique that can be employed to provide a contribution measure in the form of multiplier weight values for features based on feedback from the results of Machine Learning (ML) algorithm metrics [45]. Meanwhile, hyperparameter tuning is an attempt to adjust the parameter values of an algorithm to improve the performance of the model. The main challenges in FW and hyperparameters are the large number of parameter value candidates, the way to avoid using brute force or random origin searches, and the way to increase the speed in determining them [46].

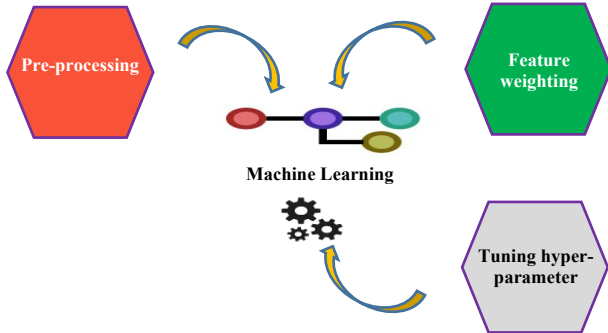


Fig. 1. Pre-processing, Feature weighting, and tuning hyperparameter

### B. Propose Machine learning methods of TVIWACRI-PSO-ELM

This research proposes a combination of the TVIWACRI-PSO algorithm, where Inertia Weight (IW) based on Eq. 1 and Acceleration Coefficients (AC) based on Eq. 3-4 for meta-heuristic determination of FW and Hyperparameter (HP) with ELM classifier on peptide datasets from Kaggle [47] in the form of b-cell, sars, and sars-cov-2 data [48]. PSO is expected to assist and facilitate the search for FW-HP values because it is very difficult to do manually due to the infinite combinations of values. When attempted without optimization techniques, it takes too much time and is not feasible to carry out in order to plan an ideal test system to achieve better

results. Manual efforts can get stuck in bias conditions when viewing the continuous values of search variables, the area of which is unlimited or includes very broad and deep expansions.

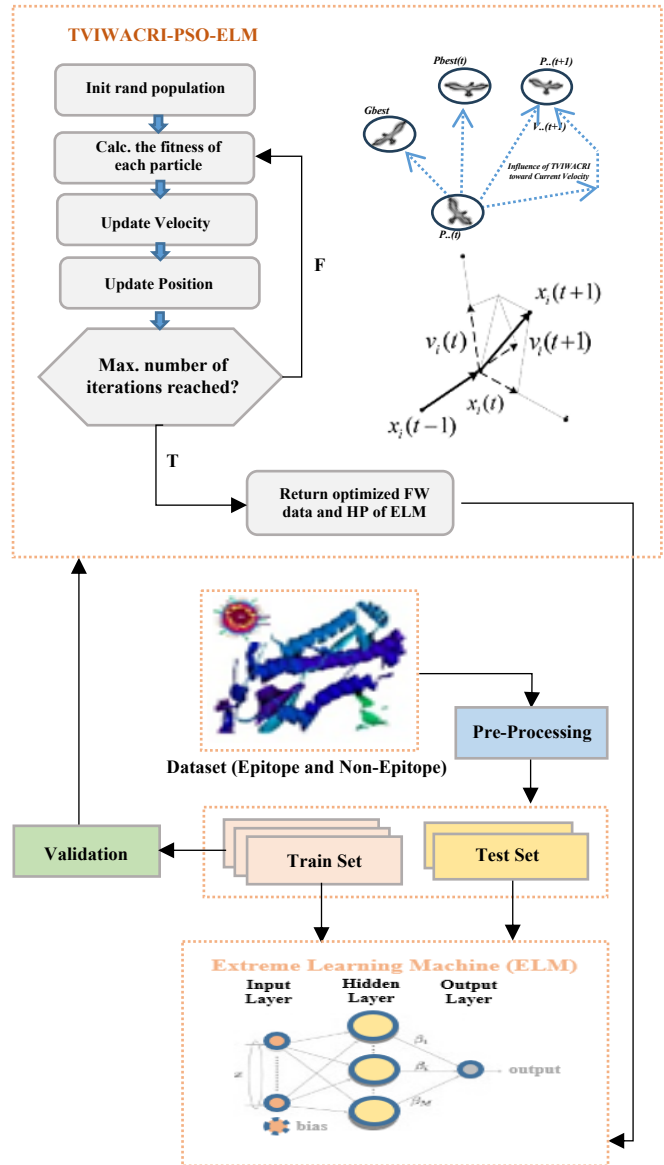


Fig. 2. ELM Optimization Scheme using TVIWACRI-PSO-ELM

Fig. 3 displays the form of PSO particle representation in this study presented in the form of vectors. Each vector cluster is marked with cluster 1 and cluster 2. The 1st cluster is for the optimization of feature weight data and the 2nd cluster is for the optimization of input weight and bias of the ELM algorithm with the fitness value equivalent to the global accuracy value as in Eq. 9. The reason why this optimization is limited to these two parts is because they are the main parts for the key to obtain better evaluation results compared to performing without the optimization process.

| Cluster 1<br>for Opt. FW |        |    |        | Cluster 2<br>for Opt. Input Weight and biases<br>ELM |    |            |    |
|--------------------------|--------|----|--------|------------------------------------------------------|----|------------|----|
| $wf_1$                   | $wf_2$ | .. | $wf_n$ | $w\_in_{1..j,k}$                                     | .. | $b_{1..j}$ | .. |

Fig. 3. Form of particle structure representation on TVIWACRI-PSO-ELM

Concretely, at the initial iteration ( $t = 0$ ), each velocity value is filled with a zero value, and the initial position utilize Eq. 1. Updating the learning velocity ( $v$ ) of particle position ( $x$ ) with several scenarios to be tested to as solution candidate. Choosing to utilize or combining them in Eq. 1-6. Random Injection (RI) process will be carried out, if **mod**( $iter, g$ ) is equal zero of current iterations, namely by re-generating, e.g. 20% of the individual particles that have the lowest fitness. This RI process is to prevent premature convergence events. Pbest is taken based on the best fitness value of each personal memory, while Gbest is taken only one particle from the overall particles or best of the best from Pbest.

$$x = x_{min} + rand[0; 1](x_{max} - x_{min}) \quad (1)$$

$$w = w_{min} + w_{max}(w_{max} - w_{min}) \frac{(t_{max} - t)}{t_{max}} \quad (2)$$

$$c_1 = (c_{1f} - c_{1i}) \frac{t}{t_{max}} + c_{1i} \quad (3)$$

$$c_2 = (c_{2f} - c_{2i}) \frac{t}{t_{max}} + c_{2i} \quad (4)$$

$$v_{i,j}^{t+1} = w \cdot v_{i,j}^t + c_1 \cdot r_1 (Pbest_{i,j}^t - x_{i,j}^t) + c_2 \cdot r_2 (Gbest_{g,j}^t - x_{i,j}^t) \quad (5)$$

$$x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^{t+1} \quad (6)$$

### C. Performance measurement for value of evaluation

Evaluation metrics are utilized as a performance measure from the results of the ML algorithm trials. The form of component variables from the values of accuracy, precision, recall, f1-score and auc can be calculated using values from TP, FP, FN, and TN [49], [50]. There are several choices of techniques that can be used, i.e., micro (for balanced data), macro (for imbalanced data) and weighted (for imbalanced data by giving weight values according to the amount of data for each class). Dissertation average is also a technique to condition the division value so that it is not zero, resulting in the undefined value. The evaluation value must also pay attention to suitability with the data processed in context because there are many options for measuring it, for example for processing data that results in generating documents that are more complex than just outputting binary class or multi-class labels [51]–[53].

|              |                 | Class by System |                 |
|--------------|-----------------|-----------------|-----------------|
|              |                 | Epitope (+)     | Non-Epitope (-) |
| Actual Class | Epitope (+)     | TP              | FN              |
|              | Non-Epitope (-) | FP              | TN              |

Fig. 4. Mapping Performance measurement from contingency table

$$Precision (P) = TP / (TP + FP) \quad (7)$$

$$Recall (R) = TP / (TP + FN) \quad (8)$$

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (9)$$

$$F1-Score = 2PR / (P + R) \quad (10)$$

In particular, Eq. 8 can be utilized as the value of the calculation results for the specific accuracy of the positive class or epitope class, and when viewed from the overall condition, then based on Fig. 4, Eq. 8 includes the formula for Recall. However, in general, the accuracy formula can be used by involving the entire classes. The opt of using the equation can be adjusted to the needs of the analysis of the test results and can be very flexible to make variations in the presentation of the results that facilitate the analysis.

## III. EXPERIMENTAL RESULTS

The test scenario was carried out by comparing the results of the ELM, RF and TVIWACRI-PSO-ELM algorithms with FW and HP optimization. RF was involved because it was going to be compared with the previous studies, in which RF had the highest results in terms of accuracy compared to the Decision Tree (DT), K-Nearest Neighbor (KNN), Logistic Regression, and Support Vector Machine (SVM). The results of the previous test were used as a basis for the continuation of the next tests. Each test utilized several parameter changes to the PSO and ELM algorithms. PSO parameters included  $wmin$ ,  $wmax$ ,  $c1f$ ,  $c2i$ ,  $c2f$ , and  $IterMax$ . Meanwhile, some of the parameters in ELM included the number of hidden nodes, random interval limits of the weight [-1;1], bias [-0.5;0.5], and the type of activation function in the form of a sigmoid, as well as a comparison of training and test data by 80% compared to 20%.

### A. Comparison of metrics values

The ELM and RF results show that ELM is better. Therefore, ELM was selected for optimization using TVIWACRI-PSO-ELM without optimizing RF. RF is used due to the reference from the previous research but no RF optimization process was carried out because the original method does not have a higher accuracy value than ELM. The results of the improvement rate of the ELM algorithm with TVIWACRI-PSO-ELM on the b-cell and SARS dataset from Table I were 0.005 and 0.163, respectively.

TABLE I. EXPERIMENT RESULT BETWEEN DATASET

| Model            | B-Cell       |                                                                                                                                                                                                       | SARS         |                                                                                                                                                  |
|------------------|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
|                  | Acc.         | Items                                                                                                                                                                                                 | Acc.         | Items                                                                                                                                            |
| TVIWACRI-PSO-ELM | <b>0.737</b> | <b>FW:</b><br>len peptide (0.01),<br>chou fasman (0.13),<br>emini (0.12),<br>kolaskar tongaonkar (0.34),<br>parker (0.58),<br>isoelectric point (0.96),<br>hydrophobicity (0.69),<br>stability (0.78) | <b>0.913</b> | <b>FW:</b><br>len peptide (1.0),<br>chou fasman (0.96),<br>parker (0.02),<br>isoelectric point (1.0),<br>aromaticity (0.96),<br>stability (0.28) |
| ELM (Original)   | 0.732        | $af = \text{sigmoid}, h = 18$                                                                                                                                                                         | 0.750        | $af = \text{sigmoid}, h = 18$                                                                                                                    |
| RF (Original)    | 0.712 (+)    | from previous reference                                                                                                                                                                               | 0.565 (+)    | from previous reference                                                                                                                          |

The increase in the value of the b-cell has been obtained, but it was not that big, while the increase in the value of the

SARS was quite high. According to the improvement in the increase in the accuracy value, it shows that the automation of FW and HP searches is very significant in facilitating the search for the right parameter values for the data classification process. The main thing that makes PSO-ELM tend to always have a great chance of obtaining a better solution than ELM is because the search technique is not just random, but organized regularly by meta-heuristic techniques during the search process. The technique is also equipped with a stage of filtering the results of the previous solution as a consideration for the colony of particle swarms to advance and move in the next iteration which is stored in the Pbest and Gbest variables. The objective function used really facilitates the success measurement of each solution candidate obtained.

### B. Analysis of convergence condition

In the convergence condition with a maximum iteration of 250 for b-cell and 250 also for SARS, the TVIWACRI-PSO-ELM algorithm displayed a kind of gradual steps in improving the solution in each iteration. Fig. 4 shows that a stable value was obtained starting from the 166th iteration for b-cell and 177th iteration for SARS. This stability means that the particle collection has reached the condition with the highest fitness value. This uniqueness is the advantage of the PSO algorithm which has the ability to learn in order to obtain a better or at least the same iteration process than the previous one.

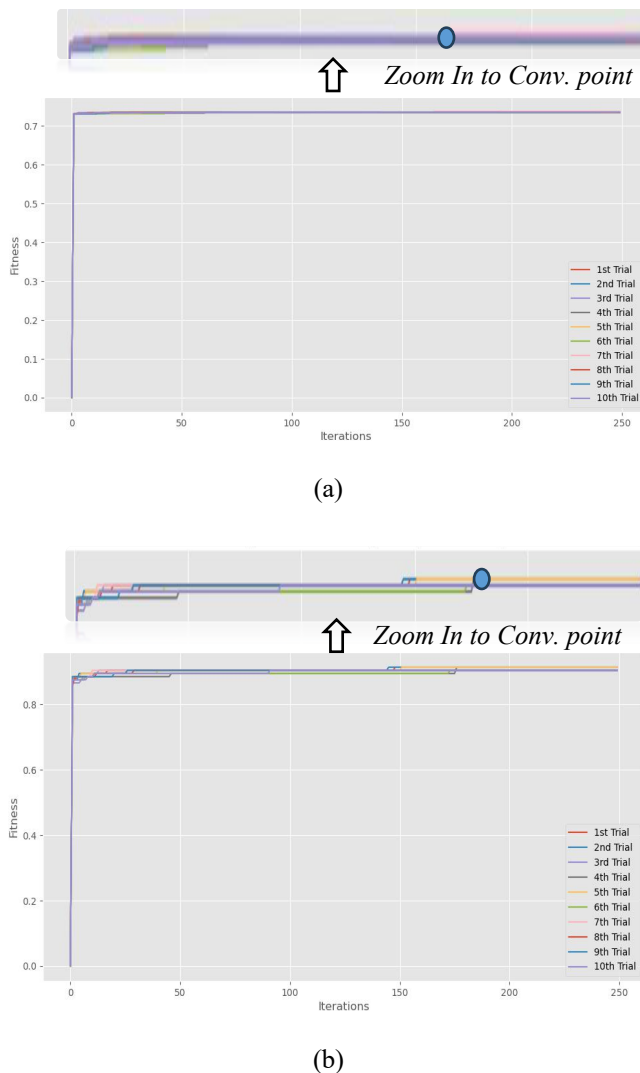


Fig. 5. (a) b-cell epitope convergence testing and (b) sars epitopes, using a combination of TVIWACRI-PSO-ELM

## IV. CONCLUSION

The TVIWACRI-PSO-ELM algorithm has been successfully implemented for b-cell and SARS epitope classification cases. The stages started from optimizing the weights on the dataset features and were concurrently with optimizing the input weights in the ELM algorithm. The essence of the basic optimization is easy automation without guessing manually which takes extra effort rather than prioritizing smart strategies to avoid a kind of brute force scheme trap. This smart strategy is a meta-heuristics technique in the PSO algorithm. The comparison results of the TVIWACRI-PSO-ELM test proved to be superior and able to increase the average accuracy by 8.4% in all datasets that were carried out in the classification process.

### CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Author 1: Collecting data and ideas, writing a framework for the concept of contribution, implementation, analysis, conclusions, and suggestions. Author 2: Mapping the research and novelty, writing corrections and scientific justification for contributions. Author 3: Comparing many ML, motivating to try test scenarios. Author 4: Determining data variable, contribution level, and publication target. Author 5: Consulting Bio-Molecular Expert, checking explanation of medical terms, performing data validation.

### ACKNOWLEDGMENT

We would like to express our gratitude to the Department of Informatics Engineering, the Faculty of Intelligent Electrical and Informatics Technology ITS and the Faculty of Medicine and the Faculty of Computer Science UB for facilitating the collaboration and research location as well as my Doctoral permit, and especially to the Bio-Molecular expert, Dr. apt. Valentina Yurina.

### REFERENCES

- [1] A. Joshi, B. C. Joshi, M. A. Mannan, and V. Kaushik, "Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach," *Inform. Med. Unlocked*, vol. 19, p. 100338, Jan. 2020, doi: 10.1016/j.imu.2020.100338.
- [2] R. Jain, A. Jain, and S. kumar Verma, "Prediction of Epitope based Peptides for Vaccine Development from Complete Proteome of Novel Corona Virus (SARS-COV-2) Using Immunoinformatics," *Int. J. Pept. Res. Ther.*, vol. 27, no. 3, pp. 1729–1740, 2021, doi: 10.1007/s10989-021-10205-z.
- [3] J. Steere-Williams, "Endemic fatalism and why it will not resolve COVID-19," *Public Health*, vol. 206, pp. 29–30, May 2022, doi: 10.1016/j.puhe.2022.02.011.
- [4] E. B. Are, Y. Song, J. E. Stockdale, P. Tupper, and C. Colijn, "COVID-19 endgame: From pandemic to endemic? Vaccination, reopening and evolution in low- and high-vaccinated populations," *J. Theor. Biol.*, vol. 559, p. 111368, Feb. 2023, doi: 10.1016/j.jtbi.2022.111368.
- [5] R. Mahajan, W. M. Lim, S. Kumar, and M. Sareen, "COVID-19 and management education: From pandemic to endemic," *Int. J. Manag. Educ.*, vol. 21, no. 2, p. 100801, Jul. 2023, doi: 10.1016/j.ijme.2023.100801.
- [6] A. Celani and P. Giudici, "Endemic–epidemic models to understand COVID-19 spatio-temporal evolution,"

- Spat. Stat.*, vol. 49, p. 100528, Jun. 2022, doi: 10.1016/j.spasta.2021.100528.
- [7] M.-H. Lee and B.-J. Kim, "COVID-19 vaccine development based on recombinant viral and bacterial vector systems: combinatorial effect of adaptive and trained immunity," *J. Microbiol. Seoul Korea*, vol. 60, no. 3, pp. 321–334, Mar. 2022, doi: 10.1007/s12275-022-1621-2.
  - [8] S. Afkhami *et al.*, "Respiratory mucosal delivery of next-generation COVID-19 vaccine provides robust protection against both ancestral and variant strains of SARS-CoV-2," *Cell*, vol. 185, no. 5, pp. 896–915.e19, Mar. 2022, doi: 10.1016/j.cell.2022.02.005.
  - [9] W.-R. Luo *et al.*, "Novel coronavirus mutations: Vaccine development and challenges," *Microb. Pathog.*, vol. 173, p. 105828, Dec. 2022, doi: 10.1016/j.micpath.2022.105828.
  - [10] N. Krishnan, L. J. Kubiawicz, M. Holay, J. Zhou, R. H. Fang, and L. Zhang, "Bacterial membrane vesicles for vaccine applications," *Adv. Drug Deliv. Rev.*, vol. 185, p. 114294, Jun. 2022, doi: 10.1016/j.addr.2022.114294.
  - [11] E. Molteni *et al.*, "Post-vaccination infection rates and modification of COVID-19 symptoms in vaccinated UK school-aged children and adolescents: A prospective longitudinal cohort study," *Lancet Reg. Health - Eur.*, vol. 19, p. 100429, Aug. 2022, doi: 10.1016/j.lanepe.2022.100429.
  - [12] A. A. Chrissian *et al.*, "Impact of COVID-19 vaccine-associated side effects on health care worker absenteeism and future booster vaccination," *Vaccine*, vol. 40, no. 23, pp. 3174–3181, May 2022, doi: 10.1016/j.vaccine.2022.04.046.
  - [13] Y. A. M. Hassan *et al.*, "A retrospective evaluation of side-effects associated with the booster dose of Pfizer-BioNTech/BNT162b2 COVID-19 vaccine among females in Eastern Province, Saudi Arabia," *Vaccine*, vol. 40, no. 49, pp. 7087–7096, Nov. 2022, doi: 10.1016/j.vaccine.2022.10.009.
  - [14] C. Menni *et al.*, "Vaccine side-effects and SARS-CoV-2 infection after vaccination in users of the COVID Symptom Study app in the UK: a prospective observational study," *Lancet Infect. Dis.*, vol. 21, no. 7, pp. 939–949, Jul. 2021, doi: 10.1016/S1473-3099(21)00224-3.
  - [15] A. Tavilani, E. Abbasi, F. Kian Ara, A. Darini, and Z. Asefy, "COVID-19 vaccines: Current evidence and considerations," *Metab. Open*, vol. 12, p. 100124, Dec. 2021, doi: 10.1016/j.metop.2021.100124.
  - [16] X. Sun, Y. Liu, M. Xu, H. Chen, J. Han, and K. Wang, "Feature selection using dynamic weights for classification," *Knowl.-Based Syst.*, vol. 37, pp. 541–549, Jan. 2013, doi: 10.1016/j.knosys.2012.10.001.
  - [17] H. Zhou, X. Li, C. Wang, and Y. Ma, "A feature selection method based on term frequency difference and positive weighting factor," *Data Knowl. Eng.*, vol. 141, p. 102060, Sep. 2022, doi: 10.1016/j.datak.2022.102060.
  - [18] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1060–1073, Apr. 2022, doi: 10.1016/j.jksuci.2019.06.012.
  - [19] S. Zhao, M. Wang, S. Ma, and Q. Cui, "A feature selection method via relevant-redundant weight," *Expert Syst. Appl.*, vol. 207, p. 117923, Nov. 2022, doi: 10.1016/j.eswa.2022.117923.
  - [20] L. Morán-Fernández, V. Bólon-Canedo, and A. Alonso-Betanzos, "How important is data quality? Best classifiers vs best features," *Neurocomputing*, vol. 470, pp. 365–375, Jan. 2022, doi: 10.1016/j.neucom.2021.05.107.
  - [21] Z. Wang, C. Wang, J. Wei, and J. Liu, "Multi-class feature selection by exploring reliable class correlation," *Knowl.-Based Syst.*, vol. 230, p. 107377, Oct. 2021, doi: 10.1016/j.knosys.2021.107377.
  - [22] L. Wang, S. Jiang, and S. Jiang, "A feature selection method via analysis of relevance, redundancy, and interaction," *Expert Syst. Appl.*, vol. 183, p. 115365, Nov. 2021, doi: 10.1016/j.eswa.2021.115365.
  - [23] M. S. Pathan, A. Nag, M. M. Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction," *Healthc. Anal.*, vol. 2, p. 100060, Nov. 2022, doi: 10.1016/j.health.2022.100060.
  - [24] I. Jain, V. K. Jain, and R. Jain, "Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification," *Appl. Soft Comput.*, vol. 62, pp. 203–215, Jan. 2018, doi: 10.1016/j.asoc.2017.09.038.
  - [25] Y. Wang, X. Jiang, F. Yan, Y. Cai, and S. Liao, "The GRA-two algorithm for massive-scale feature selection problem in power system scenario classification and prediction," *Energy Rep.*, vol. 7, pp. 293–303, Apr. 2021, doi: 10.1016/j.egyr.2021.01.067.
  - [26] J. Zhang, Y. Li, W. Xiao, and Z. Zhang, "Non-iterative and Fast Deep Learning: Multilayer Extreme Learning Machines," *J. Frankl. Inst.*, vol. 357, no. 13, pp. 8925–8955, Sep. 2020, doi: 10.1016/j.jfranklin.2020.04.033.
  - [27] L. Zha, K. Ma, G. Li, J. Yang, and Q. Fang, "An improved extreme learning machine with self-recurrent hidden layer," *Adv. Eng. Inform.*, vol. 54, p. 101736, Oct. 2022, doi: 10.1016/j.aei.2022.101736.
  - [28] S. M. Naik, R. P. K. Jagannath, and V. Kuppili, "Fractional Tikhonov regularization to improve the performance of extreme learning machines," *Phys. Stat. Mech. Its Appl.*, vol. 551, p. 124034, Aug. 2020, doi: 10.1016/j.physa.2019.124034.
  - [29] I. P. D. Wibawa, C. Machbub, A. S. Rohman, and E. Hidayat, "Reduced Computation for Extreme Learning Machine Based on Cholesky Factorization," in *2022 13th Asian Control Conference (ASCC)*, May 2022, pp. 488–493, doi: 10.23919/ASCC56756.2022.9828328.
  - [30] W. Cai, J. Yang, Y. Yu, Y. Song, T. Zhou, and J. Qin, "PSO-ELM: A Hybrid Learning Model for Short-Term Traffic Flow Forecasting," *IEEE Access*, vol. 8, pp. 6505–6514, 2020, doi: 10.1109/ACCESS.2019.2963784.
  - [31] M. R. Kaloop *et al.*, "Particle Swarm Optimization Algorithm-Extreme Learning Machine (PSO-ELM) Model for Predicting Resilient Modulus of Stabilized Aggregate Bases," *Appl. Sci.*, vol. 9, no. 16, Art. no. 16, Jan. 2019, doi: 10.3390/app9163221.
  - [32] X.-L. Li, R. Serra, and J. Olivier, "A multi-component PSO algorithm with leader learning mechanism for

- structural damage detection,” *Appl. Soft Comput.*, vol. 116, p. 108315, Feb. 2022, doi: 10.1016/j.asoc.2021.108315.
- [33] H. Xue, Y. Bai, H. Hu, T. Xu, and H. Liang, “A Novel Hybrid Model Based on TVIW-PSO-GSA Algorithm and Support Vector Machine for Classification Problems,” *IEEE Access*, vol. 7, pp. 27789–27801, 2019, doi: 10.1109/ACCESS.2019.2897644.
- [34] X. Xu, J. Li, and H. Chen, “Enhanced support vector machine using parallel particle swarm optimization,” in *2014 10th International Conference on Natural Computation (ICNC)*, Aug. 2014, pp. 41–46. doi: 10.1109/ICNC.2014.6975807.
- [35] M. Rostami, S. Forouzandeh, K. Berahmand, and M. Soltani, “Integration of multi-objective PSO based feature selection and node centrality for medical datasets,” *Genomics*, vol. 112, no. 6, pp. 4370–4384, Nov. 2020, doi: 10.1016/j.ygeno.2020.07.027.
- [36] M. Haouari and M. Mhiri, “A particle swarm optimization approach for predicting the number of COVID-19 deaths,” *Sci. Rep.*, vol. 11, p. 16587, Aug. 2021, doi: 10.1038/s41598-021-96057-5.
- [37] P. Cihan and Z. B. Ozger, “A new approach for determining SARS-CoV-2 epitopes using machine learning-based in silico methods,” *Comput. Biol. Chem.*, vol. 98, p. 107688, Jun. 2022, doi: 10.1016/j.compbiolchem.2022.107688.
- [38] K. Sundararajan and K. Srinivasan, “Feature-Weighting-Based Prediction of Drought Occurrence via Two-Stage Particle Swarm Optimization,” *Sustainability*, vol. 15, no. 2, Art. no. 2, Jan. 2023, doi: 10.3390/su15020929.
- [39] M. Kashina, I. D. Lenivtceva, and G. D. Kopanitsa, “Preprocessing of unstructured medical data: the impact of each preprocessing stage on classification,” *Procedia Comput. Sci.*, vol. 178, pp. 284–290, Jan. 2020, doi: 10.1016/j.procs.2020.11.030.
- [40] K. Maharana, S. Mondal, and B. Nemade, “A review: Data pre-processing and data augmentation techniques,” *Glob. Transit. Proc.*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.
- [41] H. S. Obaid, S. A. Dheyab, and S. S. Sabry, “The Impact of Data Pre-Processing Techniques and Dimensionality Reduction on the Accuracy of Machine Learning,” in *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*, Mar. 2019, pp. 279–283. doi: 10.1109/IEMECONX.2019.8877011.
- [42] A. Sarker, Md. R. Islam, and A. Y. Srizon, “A Comprehensive Pre-processing Approach for High-Performance Classification of Twitter Data with several Machine Learning Algorithms,” in *2020 IEEE Region 10 Symposium (TENSYPMP)*, Jun. 2020, pp. 630–633. doi: 10.1109/TENSYPMP50017.2020.9230590.
- [43] A. K. B and M. M. Kodabagi, “Efficient Data Preprocessing approach for Imbalanced Data in Email Classification System,” in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, Oct. 2020, pp. 338–341. doi: 10.1109/ICSTCEE49637.2020.9277221.
- [44] M. P. Singh, V. Gayathri, and D. Chaudhuri, “A Simple Data Preprocessing and Postprocessing Techniques for SVM Classifier of Remote Sensing Multispectral Image Classification,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, pp. 7248–7262, 2022, doi: 10.1109/JSTARS.2022.3201273.
- [45] I. Niño-Adan, D. Manjarres, I. Landa-Torres, and E. Portillo, “Feature weighting methods: A review,” *Expert Syst. Appl.*, vol. 184, p. 115424, Dec. 2021, doi: 10.1016/j.eswa.2021.115424.
- [46] S. Nematzadeh, F. Kiani, M. Torkamanian-Afshar, and N. Aydin, “Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases,” *Comput. Biol. Chem.*, vol. 97, p. 107619, Apr. 2022, doi: 10.1016/j.compbiolchem.2021.107619.
- [47] F. Corporation, V. Basu, and M. Prata, “COVID-19/SARS B-cell Epitope Prediction,” 2021. <https://www.kaggle.com/datasets/futurecorporation/epitope-prediction> (accessed May 23, 2023).
- [48] J. Singh, D. Malik, and A. Raina, “Immuno-informatics approach for B-cell and T-cell epitope based peptide vaccine design against novel COVID-19 virus,” *Vaccine*, vol. 39, no. 7, pp. 1087–1095, Feb. 2021, doi: 10.1016/j.vaccine.2021.01.011.
- [49] J. Lever, M. Krzywinski, and N. Altman, “Classification evaluation,” *Nat. Methods*, vol. 13, no. 8, Art. no. 8, Aug. 2016, doi: 10.1038/nmeth.3945.
- [50] X. Zhou and A. Del Valle, “Range Based Confusion Matrix for Imbalanced Time Series Classification,” in *2020 6th Conference on Data Science and Machine Learning Applications (CDMA)*, Mar. 2020, pp. 1–6. doi: 10.1109/CDMA47397.2020.00006.
- [51] S. Lee et al., “A Survey on Evaluation Metrics for Machine Translation,” *Mathematics*, vol. 11, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/math11041006.
- [52] A. Tasnim, N. Hossain, N. Parvin, S. Tabassum, R. Rahman, and M. Iqbal Hossain, “Experimental Analysis of Classification for Different Internet of Things (IoT) Network Attacks Using Machine Learning and Deep learning,” in *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, Mar. 2022, pp. 406–410. doi: 10.1109/DASA54658.2022.9765108.
- [53] A. Diallo, L. Affognon, C. Diallo, and E. C. Ezin, “Deep Learning Based Binary and Multi-class Classification Comparison for Anomaly Detection,” in *2022 International Conference on Engineering and Emerging Technologies (ICEET)*, Oct. 2022, pp. 1–6. doi: 10.1109/ICEET56468.2022.10007171.