

HML-PCD: A Hybrid machine learning technique for Early prediction and classification of celiac disease

Mayura Tapkire*¹, Vanishri Arun², Lavanya M. S.³

Submitted: 25/01/2024 Revised: 03/03/2024 Accepted: 11/03/2024

Abstract: This research proposes a hybrid machine learning approach for early celiac disease (CD) prediction. The method incorporates an enhanced cuttlefish optimization (ECO) algorithm for data pre-processing, removing unwanted artefacts. It employs an improved whale optimization algorithm (IWO) to extract multi-class features like texture and non-linear feature vectors for efficient disease detection. Subsequently, a modified crow search (MCS) algorithm optimizes feature selection, addressing dimensionality issues. Finally, a hybrid probabilistic deep neural network (PDNN) is introduced for CD prediction/classification, enhancing detection accuracy. The proposed approach demonstrates higher accuracy, precision, recall, and F-measure compared to current best practices.

Keywords: Celiac disease (CD), data pre-processing, feature extraction, feature selection, CD prediction/classification

1. Introduction

Celiac Disease (CD), an inflammatory reaction to gluten, damages the small intestine mucosa and poses various health risks [1][2][3][4]. Diagnosis involves testing for tissue transglutaminase antibodies, with endoscopy as the gold standard [5][6]. The challenge lies in detecting distal small intestine abnormalities noninvasively [7]. CD, prevalent in Western countries, affects genetically predisposed individuals, with many cases going undetected [8][9]. Autoimmune disorders (AD), including digestive illnesses like Crohn's disease, are on the rise [10][11][12][13]. CD, triggered by dietary gluten in genetically predisposed individuals, involves auto-antibodies damaging the digestive tract mucosa [14][15]. Symptoms vary, and diagnosis relies on intestinal biopsies and serological alterations [16][17]. Case-finding is deemed ineffective, prompting the need for a machine learning model to detect undiagnosed CD [18][19][20].

2. Related Work

Researchers have employed various methods for celiac disease detection and diagnosis. An AuNP-enhanced GQD/PAMAM hybrid [21], utilizing EDC-NHS cross-linking chemistry, was used to detect celiac disease. An immunosensor by Rosales et al. [22] detected anti-tissue transglutaminase antibodies, incorporating chemically added disulfide groups for IgA and IgG autoantibody

detection. Smarrazzo et al. [23] explored ESPGHAN guidelines in diagnosing CD in children across 13 Mediterranean countries. Erlichster et al. [24] proposed CD-LAMP for genotyping HLA-DQ2.5, HLA-DQ8, HLA-DQ2.2, and HLA-DQA1*05 with high accuracy and speed. Koh et al. [25, 26, 27, 28] introduced a CAD system using DWT for nonlinear and textural information from movies. Wang et al. [29] proposed deep learning calibration modules for celiac disease detection, integrating ResNet50 and Inception-v3 for image analysis. Gheshlagh et al. [30] conducted a meta-analysis on Iranian individuals with type 1 diabetes to assess CD prevalence. Russo et al. [31] introduced a method for diagnosing and treating celiac disease in children through quality of life ratings. Ralbovsky et al [32] developed a non-invasive, low-cost method for CD diagnosis using Raman hyper spectroscopy and PLS-DA for Raman spectra data. Wei Koh et al. [33] used modified Marsh ratings and machine learning, creating sub-bands with the SPT method and calculating entropy and nonlinear properties for celiac disease diagnosis.

3. Problem Methodology and System Design

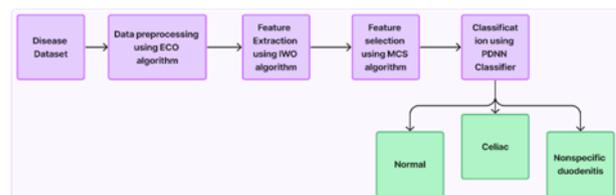


Fig 1 System design of proposed technique

Six classifiers were used to categorize features into two groups with varying accuracies. H&E-stained biopsy

¹ Research Scholar, Vellore Institute of Technology, Vellore, TN, INDIA

² Associate Professor, Vellore Institute of Technology, Vellore, TN, INDIA

* Corresponding Author Email: mvazralu1206@gmail.com

images achieved 88.89%, RGB stained images 82.92%, and multi-class images 72%. The technology holds promise for quicker and more reliable tumour diagnosis. Celiac Disease (CD), affecting 1% of the American population, is diagnosed through endoscopic biopsy, but misdiagnosis is common due to ambiguous features. New methods, including the hybrid machine learning approach (HML-PCD), improve CD identification on biopsy images, enhancing early prediction and prognosis. Figure 1 illustrates the system model and its contribution.

4. Proposed Methodology

This section explains how data pre-processing, feature extraction, and optimum feature selection operation. Then we explain disease categorization and prediction.

4.1 Data pre-processing using enhanced cuttlefish optimization algorithm

The cuttlefish optimization technique addresses numerical global optimization problems by mimicking the colour-shifting behaviour of cuttlefish. The cuttlefish's unique colours and patterns are generated by light reflecting from three layers of cells. CFO, a bio-inspired optimization method, models the cuttlefish's colour-changing behaviour using evolutionary algorithms from artificial intelligence. The enhanced cuttlefish optimization (ECO) method, an improvement upon the classic CFO, is introduced. Initially, two parameters (coded as R_1 and R_2) are used to construct the reflection strategy. Whereas the other two (coded as U_1 and U_2) are used for visibility strategy. At start, a random population (Q) is created using,

$$Q[y_j].point s[i] = random * (Upper limit - Lower limit) + Lower limit \quad (1)$$

Any number between 0 and 1 could be chosen at random. Each solution has two values, as illustrated in Fig. 3. Points and a fitness value are included in a vector of continuous values known as a points and where y_j represents optimal best solution with the current iteration.

$$H_{size} = \frac{Q}{4} \quad (2)$$

where H_{size} is the size of each group, and Q is the initial population size. In the decision-making phase, a new solution y_{new} is created using (3) until the final criteria are met.

$$y_{new} = reflection + visibilty \quad (3)$$

Where y_{new} is the new solution, this is the sum of the calculated reflections visibility. Prior to the computation of the and necessary to calculate R and u . R is set to 1 in cases 3, 4, and 5, whereas u is set to 1 in cases 1 and 2.

$$r = random() * (R_1 - R_2) + R_2 \quad (4)$$

$$u = random() * (U_1 - U_2) + U_2 \quad (5)$$

where $random()$ denotes a random number from 0 to 1. R_1 , R_2 , U_1 and U_2 manually set the settings at startup. The best result is based on the average score of Bu_{best} (6). The total score of Bu_{best} is the best solution is calculated by dividing by N , where N is the total score which reflects sum of the calculated reflections visibility.

$$Bu_{best} = \frac{\sum [bestsolution.poibts [i]]}{N} \quad (6)$$

To begin, the four search algorithms are performed to the subsets of data shown below. Using (7) and (8) to determine new solution x_{new} is created. H_1 acts as a global search, and the value of r is calculated using (4):

$$reflection = r * H_1[y_i].point s[i] \quad (7)$$

$$visibilty = u * (bestsolution.point s[i] - H_1[y_j].point s[i]) \quad (8)$$

Where H_1 is the first set of solutions, y_j is the second solution in the group. $[y_j].points [i]$ shows point i of the solution. Under the best solution, a new solution was developed using (8) and (9) to calculate the reflection and visibility accordingly. H_2 acts as a local search and calculate optimal solution using the value u (5).

$$reflection = r * bestsolution.point s[i] \quad (9)$$

Algorithm 1 Data pre-processing using ECO algorithm	
Input	:q points and y_j
Output	: H_{5-size} , Bu_{best}
1	Initialize the population Q
2	Initialize parameters R_1, R_2 and U_1, U_2
3	Calculate fitness $H_{size} = \frac{Q}{4}$
4	Calculate best solution in $Bu_{best} = \frac{\sum [bestsolution.poibts [i]]}{N}$
5	If $i=0, j=1$

6	Determine (y) vaules $y_{new} = reflection + visibility$
7	Calculate r and u $r = random() * (R_1 - R_2) + R_2$ $u = random() * (U_1 - U_2) + U_2$
8	Update position using activation functions $visibility = u * (bestsolution.point s[i] - Bu_{best})$
9	End while
1	End
0	

A new updated solution y_{new} is generated based on AVb . Equations (9) and (10) used to calculate the *reflection* and *visibility*, respectively. H_3 acts as a local search (exploitation), and the value of u is computed using (5):

$$visibility = u * (bestsolution.point s[i] - Bu_{best}) \quad (10)$$

For example, if the best solution has two points (6, -2), U_1 and U_2 are set to (1,-1), respectively. The result of the Bu_{best} will be calculated as follow;

$$Bu_{best} = (6 - 2) / 2 = 2$$

$$H_{5-size} = \left[\left((H5T_{percent} * H_{size}) * 4 \right) + \frac{H_{size}}{2} \right] \quad (11)$$

The results change from the previous four groups, only the best solutions are selected H_{5-size} as follows:

$$H_5 = (\uparrow H_1 * H5T_{percent}) + (\uparrow H_2 * H5T_{percent}) + (\uparrow H_3) \quad (12)$$

where \uparrow represents the sorted solutions from best to worst. Algorithm 1 describes the working function of enhanced cuttlefish optimization (ECO) algorithm.

4.2 Multi-level feature extraction using improved whale optimization algorithm

High-level computer vision applications depend on feature extraction. The Improved Whale Optimization method (IWO) utilizes a multi-level feature extraction process inspired by humpback whales' hunting behavior, aiming to minimize training complexity in classification techniques. The algorithm is a meta-heuristic optimization system drawing inspiration from humpback whales' bubble-net hunting approach. In IWO algorithm, we generate real numbers \vec{R}_1 and \vec{R}_2 $q \in [0,1]$ and use them to calculate

\vec{B} and \vec{D} the calculation formulas of \vec{B} and \vec{D} are as follows:

$$\vec{B} = 2 \cdot \vec{b} \cdot \vec{R}_1 - \vec{b} \quad (13)$$

$$\vec{D} = 2 \cdot \vec{R}_2 \quad (14)$$

where \vec{b} is a real number that goes linearly from 2 to 0 in the iteration. The search agent is compelled to leave its present position and wander aimlessly across space in pursuit of prey with this strategy. The function of position update is described as follows:

$$\vec{C} = \left| \vec{D} \times \vec{Q}_{rand} - \vec{Q}_L^s \right| \quad (15)$$

$$\vec{Q}_L^{s+1} = \vec{Q}_{rand} - \vec{B} \times \vec{C} \quad (16)$$

where \vec{Q}_{rand} is approximately the generated level vector on the boundary line. \vec{Q}_L^s i-th is the generation vector of the search agent status vector, \vec{Q}_L^{s+1} which is the generation of the $S + 1$ vector of the search agent status vector. The position of the prey is determined and the prey is encircled throughout this procedure. The search agent approaches the appropriate location of the search agent. The position update summarized as follows:

$$\vec{C} = \left| \vec{D} \times \vec{Q}_*^s - \vec{Q}_L^s \right| \quad (17)$$

$$\vec{Q}_L^{s+1} = \left| \vec{Q}_L^s - \vec{B} \times \vec{C} \right| \quad (18)$$

where \vec{Q}_*^s is the generation t of optimal search agent's position vector. The search agent is compelled to leave its present position and wander aimlessly across space in pursuit of prey with this strategy. The position transformation formula and mathematical model are as follows:

$$\vec{Q}_L^{s+1} = \vec{C}^s \cdot E^{aL} \cdot \cos(2\pi L) + \vec{Q}_*^s \quad (19)$$

An arbitrary real value between [-1,1] is used as the logarithmic shape constant for the logarithmic helix. In the

random stabilization phase, we use the general Cauchy's functionality of the search agent to counteract the cochlear mutation. The mathematical formula for the general function opposite the sofa is as follows:

$$f^{-1}(q; y_0, \gamma) = y_0 + \gamma \cdot \tan(\pi \cdot (q - 1/2)) \quad (20)$$

It's based on this formula that the whale optimization strategy for random prey is updated as follows:

$$\vec{Q}_L^{s+1} = \vec{q}_L^s + \vec{B} \cdot \tan(\pi \cdot (R_3 - 1/2)) \quad (21)$$

where, \vec{q}_L^s is the position of the j-th search agent in generation s.

R_3 is a random number within the interval of [0, 1]. A local mutation probability K is supplied to assure the algorithm's stability. Retain the mutant global-best search agent who is more fit. As an example, consider the following formula:

$$\vec{Q}_{*M}^s = \begin{cases} \vec{Q}_*^s \cdot (1+n) & , R_4 > K \\ \vec{Q}_*^s & ,, R_4 \leq K \end{cases} \quad (22)$$

The algorithm 2 describes the working function of multi-level feature extraction using improved whale optimization (IWO) algorithm.

Algorithm 2 Feature extraction using IWO algorithm	
Input	: \vec{B} and \vec{D}
Output	: \vec{Q}_{*M}^s
1	Generate the initial population
2	Evaluate the fitness for each candidate solutions in
3	while The halting criterion is not satisfied do
4	For j=1 to nq do
5	Update the values $\vec{Q}_L^{s+1} = \vec{Q}_{rand} - \vec{B} \times \vec{C}$
6	boundary range $\vec{Q}_L^{s+1} = \vec{C}^s \cdot E^{aL} \cdot \cos(2\pi L) + \vec{Q}_*^s$
7	formula of whale optimization algorithm $\vec{Q}_L^{s+1} = \vec{q}_L^s + \vec{B} \cdot \tan(\pi \cdot (R_3 - 1/2))$
8	Global best search agent has a better fitness $\vec{Q}_{*M}^s = \begin{cases} \vec{Q}_*^s \cdot (1+n) & , R_4 > K \\ \vec{Q}_*^s & ,, R_4 \leq K \end{cases}$
9	End

4.3 Optimal feature selection using modified crow search algorithm

We devised an improved feature selection method after extraction, halving the number of incorrectly chosen features while maintaining true positive rates. This optimizes efficiency, resulting in a simpler, more comprehensible, and accurate model. Inspired by crow behaviour, the modified crow search method (MCS) employs chaos theory in an evolutionary computing approach. Unlike traditional methods, MCS doesn't require following a specific crow (crow I) but determines and updates its location relative to another crow's (crow j) food storage.

$$p_j^{(s+1)} = p_j^{(s)} + R_j * fL_i^{(s)} * (M_i^{(s)} - p_j^{(s)}), \quad (23)$$

where f_L indicates the flight length. R_j been a random number $\in [0, 1]$. When crow I realizes this, crow j pursues her in order to locate her meal. In this case, the crow I travels at random to deceive the crow j. mathematically, the two situations may be joined as follows:

$$p_j^{(s+1)} = \begin{cases} p_j^{(s)} + R_j * fL_i^{(s)} * (M_i^{(s)} - p_j^{(s)}), & R_i \geq Bq_j^s \\ \text{choose random position,} & \text{otherwise} \end{cases} \quad (24)$$

Crows are evaluated using a defined fitness function, adjusting their positions based on the scores. Viability of new positions is assessed, and crows' memories are updated with fresh information.

$$M_j^{(s+1)} = \begin{cases} p_j^{(s)} & \text{if } f(M_j^{(s+1)}) \text{ is better than } f(p_j^{(s)}) \\ M_j^{(s+1)}, & \text{otherwise} \end{cases} \quad (25)$$

The set of solutions is converted to binary dimensional, where the solution is defined as binary dimmer 0, 1. This is done by moving the agents from the constant to the binary space

$$p_j^{(s+1)} = \begin{cases} 1 & \text{if } (t(p_j^{(s+1)})) \geq rand() \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

$$\text{wheret} = \frac{1}{1 + E^{10(p_j^{(s+1)} - 0.5)}}$$

Chaos is considered a phenomenon wherein any alteration in the system's state can lead to nonlinear changes in subsequent actions. The Crow Search Algorithm (CSA) employs random variables to diversify the crow's position in the Chaos Crow Search Algorithm (CCSA). The random

factors used for varying the crow's location in CCSA include:

$$p_j^{(s+1)} = \begin{cases} p_j^{(s)} + D_i * (M_i^{(s)} - p_j^{(s)}), & D_w \geq Bq_j^s \\ \text{choose } b \text{ random position,} & \text{otherwise} \end{cases} \quad (27)$$

Where D_i stands for value gotten from chaotic map at i -th iteration and D_w the value gotten from the chaotic map at w -th iteration.

$$U_{shape} = \left| \frac{2}{\pi} \arctan\left(\frac{\pi}{2} p_j^s\right) \right| \quad (28)$$

This has led to the development of a fitness function for discovering solutions that balances the two goals.

$$Fitness = \alpha \Delta_r(C) + \beta \frac{|X|}{|S|} \quad (29)$$

Algorithm 3 describes the working function of optimal feature selection using MCS algorithm.

Algorithm 3 Optimal feature selection using MCS algorithm	
Input	: extracted features
Output	: optimally select best features
1	Start,
2	Initialize the population of crows R_j
3	Evaluate the fitness fL $Fitness = \alpha \Delta_r(C) + \beta \frac{ X }{ S }$
4	binary form of {0, 1}
5	Update the CSA $p_j^{(s+1)} = \begin{cases} p_j^{(s)} + R_j * fL_i^{(s)} * (M_i^{(s)} - p_j^{(s)}), & R \\ \text{choose } b \text{ random position,} & \text{otherwise} \end{cases}$
6	boundary range $p_j^{(s+1)} = \begin{cases} 1 & \text{if } (t(p_j^{(s+1)})) \geq rand() \\ 0, & \text{otherwise} \end{cases}$
7	Check the feasibility of the crow $U_{shape} = \left \frac{2}{\pi} \arctan\left(\frac{\pi}{2} p_j^s\right) \right $
8	Produce the best solution $M_j^{(s+1)} = \begin{cases} p_j^{(s)} & \text{if } f(M_j^{(s+1)}) \text{ is better than } f(p_j^{(s)}) \\ M_j^{(s+1)}, & \text{otherwise} \end{cases}$
9	End

4.4 Disease prediction and classification using hybrid probabilistic deep neural network

Doctors traditionally use statistical methods and personal judgment to assess prognosis and disease risk, leading to biases, errors, and increased costs in patient care. Electronic health data enables the application of advanced computational technologies like machine learning. The hybrid probabilistic deep neural network (PDNN) improves upon deep neural networks by incorporating probabilistic layers instead of weights for enhanced probabilistic deep learning. The layer parameters for PDNN were set up as follows:

$$g_j = sig\left(a_j + \sum_i^N z_{ji} u_i\right) \quad (30)$$

$$s_i = sig\left(d_j + \sum_i^M z_{ji} g_i\right) \quad (31)$$

where $sig(\bullet)$ stands for the sigmoid function, u stands for the input space vectors as well as a desired goal vector, and s stands for the neural network's projected output space vectors. The program's goal is to reduce errors, or values that differ between visible and predicted neurons. In general, we create probabilistic functions by generating the probabilistic model's state vector, which may be represented as

$$z_K = [z_{11,K}^g \ \& \ a_{3,K}^g \ z_{1,K}^x \ \& \ z_{3,K}^x \ a_K^x]^S \quad (32)$$

Here Z_x and Z_g are the terms of the output layer and hidden layer weight networks, respectively. PDNN Three-layer recurrent neural network measurement model process and Functions are as follows:

$$z_K = F(z_{K-1}) \quad (33)$$

$$s_K = sig(z_K^S [sig(z_K u) + a] + d) \quad (34)$$

If the error is more than zero, the weight update algorithm uses the extended Kalman filter technique, which has two phases.

$$Z_{K/K-1} = Z_{K-1/K-1} + \alpha \frac{\partial e_{K-1}}{\partial z_{K-1/K-1}} \quad (35)$$

$$Q_{K/K-1} = Q_{K-1/K-1} + P_{K-1} \quad (36)$$

where e is the error price function, which is the sum of the square error operations, the α rating factor is 0.01, and P_K the process noise factor is 0, the coverage is calculated from the $Q_{K/K-1}$ latest estimates. To improve the level of predictability ($Z_{K/K}$), we need to define the Kalman gain, which is expressed as follows:

$$k_K = Q_{K/K-1} G_K^S [G_K Q_{K/K-1} G_K^S + r_K]^{-1} \quad (37)$$

where r_K is the unknown prior covariance that is given by

$$r_K = r_{K-1} + \frac{1}{K} ((u - s_K)(u - s_K)^2 - r_{K-1}) \quad (38)$$

Where t_k is the linear composition of each output node. Last wait ($Z_{K/K}$) update and related update error contacts ($Q_{K/K}$) are as follows:

$$Q_{K/K} = Q_{K/K-1} - k_K G_K^S Q_{K/K-1} \quad (39)$$

$$Z_{K/K} = Z_{K/K-1} + \eta k_K (u - s_{K-1}) \quad (40)$$

Algorithm 4 describes the working function of CD prediction and classification using hybrid probabilistic deep neural network (PDNN).

Algorithm 4 CD prediction and classification using PDNN technique	
Input : Optimal features	
Output : $Z_{K/K}, Q_{K/K}$	
1	Calculate minimisers for the PDNN constrained optimization problem
2	Initialize parameters
3	Given training set $z_K = [z_{1,K}^g \ \ a_{3,K}^g \ \ z_{1,K}^x \ \ z_{3,K}^x \ \ a_K^x]^S$
4	Calculate mathematical terminology $s_i = sig\left(d_j + \sum_i^M z_{ji} g_i\right)$
5	End for
6	For each $s_K = sig(z_K^S [sig(z_K u) + a] + d)$

7	if $u \in randn[C]$
8	Update position using activation functions $r_K = r_{K-1} + \frac{1}{K} ((u - s_K)(u - s_K)^2 - r_{K-1})$
9	End while
10	End for $Z_{K/K} = Z_{K/K-1} + \eta k_K (u - s_{K-1})$
11	End

5. Result and Discussions

We applied our HML-PCD approach to a well-known dataset to see how well it worked. The suggested HML-PCD method's accuracy, precision, recall, F1-score, G-mean, and AUC are all compared to those of presently existing state-of-the-art approaches.

5.1 Dataset description

The University of Virginia's archives provided biopsy slides from 63 Celiac Disease (CD) patients and 63 healthy individuals, with an almost equal gender and age distribution. CD patients constituted 37.7% of the participants, and 239 duodenal biopsy slides were analysed. The CD dataset included various types, each with specific training and testing sets.

Level of celiac disease	No. of patient data's for ...		
	Trainin g	Testing	Total
Type 1	4988	2137	7125
Type 2	4790	2052	6842
Type 3	5684	2436	8120
Type 4	5670	2433	8111
Total number of patient data's	21140	9058	30198

5.2 Performance metrics

When comparing existing MWCNT [21], ESPGHAN [23], CD-LAMP [24] and PLS-DA techniques, the simulation results of our suggested HML-PCD methodology are shown to be superior in accuracy, precision, recall and G-mean and AUC.

Models	Performance metrics (%)					
	Accuracy	Precision	Recall	F1 score	G-mean	AUC
MWCNT [21]	81.18	87.08	79.07	82.01	93.07	91.74

ESPGHAN [23]	89.98	77.19	69.98	78.91	91.98	89.96
CD-LAMP [24]	94.23	85.23	93.57	97.37	86.04	87.36
PPRN [25]	93.78	82.45	95.79	98.18	90.75	92.25
PLS-DA [32]	79.02	85.78	76.04	81	95.07	94.14
ANN [33]	96.78	95.98	94.39	98.78	93.04	95.76
HML-PCD	97.85	96.78	98.12	99.14	96.78	96.08

5.3 Comparative analysis

A comparison with existing models in Table 2, including MWCNT [21], ESPGHAN [23], CD-LAMP [24], PPRN [25], PLS-DA [32], and ANN [33], reveals HML-PCD's superior performance. HML-PCD outperforms these techniques in precision by 10.023% to 20.242%, recall by 19.415% to 28.679%, and F1-score by 17.279% to 20.405%

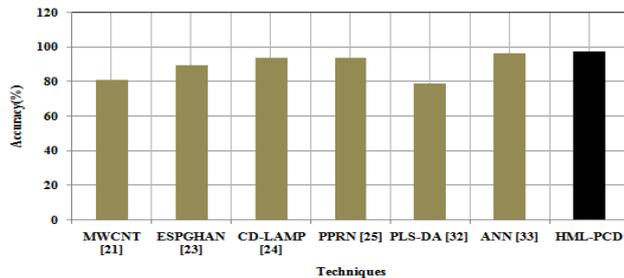


Fig. 2 Comparison of suggested and current models' accuracy

The suggested HML-PCD approach outperforms the current MWCNT [21], ESPGHAN [23], CD-LAMP [24], PPRN [25], PLS-DA [32], and ANN [33] procedures by 3.833 percent, 4.960 percent, 11.097 percent, 6.231 percent, 1.767 percent, and 3.864 percent, respectively. The AUC of proposed HML-PCD technique is 4.517%, 6.370%, 9.076%, 3.986%, 2.019% and 0.333% higher than the existing MWCNT [21], ESPGHAN [23], CD-LAMP [24], PPRN [25], PLS-DA [32] and ANN [33] techniques respectively. Figs. 2 to 7 showed the graphical representation of comparative analysis for accuracy, precision, recall, F1-score, G-mean and AUC respectively.

6. Conclusion

Celiac disease is substantially increasing in the population globally and needs techno-medical interventions [34]. According to simulation data, the proposed HML-PCD technique has a 12 percent higher average accuracy than existing state-of-the-art procedures. For the most part, the proposed HML-PCD strategy outperforms existing best practises by 13.45 percent. Compared to existing best practises, the HML-PCD methodology has an average F1 score that is 17.56 percent higher. AUC, F1score, and G-

mean are just a few of the ways HML-PCD beats current best practises in simulation.

References

- [1] Caio, Giacomo, Umberto Volta, Anna Sapone, Daniel A. Leffler, Roberto De Giorgio, Carlo Catassi, and Alessio Fasano. "Celiac disease: a comprehensive current review." *BMC medicine* 17, no. 1 (2019): 1-20.
- [2] Li, Bing Nan Nan, Xinle Wang, Rong Wang, Teng Zhou, Rongke Gao, Edward J. Ciaccio, and Peter H. Green. "Celiac disease detection from videocapsule endoscopy images using strip principal component analysis." *IEEE/ACM transactions on computational biology and bioinformatics* (2019).
- [3] Vicnesh, Jahmunah, Joel Koh En Wei, Edward J. Ciaccio, Shu Lih Oh, Govind Bhagat, Suzanne K. Lewis, Peter H. Green, and U. Rajendra Acharya. "Automated diagnosis of celiac disease by video capsule endoscopy using DAISY descriptors." *Journal of Medical Systems* 43, no. 6 (2019): 1-9.
- [4] Gulati, Shraddha, Andrew Emmanuel, Mark Ong, Polychronis Pavlidis, Mehul Patel, Tareq El-Menabaway, Zuzana Vackova et al. "Near-focus narrow-band imaging classification of villous atrophy in suspected celiac disease: development and international validation." *Gastrointestinal Endoscopy* 94, no. 6 (2021): 1071-1081.
- [5] Laserna-Mendieta, E. J., M. J. Casanova, Á. Arias, L. Arias-González, P. Majano, L. A. Mate, C. H. Gordillo-Vélez, M. Jiménez, T. Angueira, and E. Tébar-Romero. "Poor Sensitivity of Fecal Gluten Immunogenic Peptides and Serum Antibodies to Detect Duodenal Mucosal Damage in Celiac Disease Monitoring." *Nutrients* 2021, 13, 98." (2020).
- [6] Husby, Steffen, Joseph A. Murray, and David A. Katzka. "AGA clinical practice update on diagnosis and monitoring of celiac disease—changing utility of serology and histologic measures: expert review." *Gastroenterology* 156, no. 4 (2019): 885-889.

- [7] Quigley, Eamonn MM, Joseph A. Murray, and Mark Pimentel. "AGA clinical practice update on small intestinal bacterial overgrowth: expert review." *Gastroenterology* 159, no. 4 (2020): 1526-1532.
- [8] Valitutti, Francesco, and Alessio Fasano. "Breaking down barriers: how understanding celiac disease pathogenesis informed the development of novel treatments." *Digestive diseases and sciences* 64, no. 7 (2019): 1748-1758.
- [9] Mearns, Elizabeth S., Aliko Taylor, Talia Boulanger, Kelly J. Craig, Michele Gerber, Daniel A. Leffler, Jennifer Drahos, David S. Sanders, and Benjamin Lebwohl. "Systematic literature review of the economic burden of celiac disease." *Pharmacoeconomics* 37, no. 1 (2019): 45-61.
- [10] McAllister, Brian P., Emmanuelle Williams, and Kofi Clarke. "A comprehensive review of celiac disease/gluten-sensitive enteropathies." *Clinical reviews in allergy & immunology* 57, no. 2 (2019): 226-243.
- [11] Norström, Fredrik, Anna Myléus, Katrina Nordyke, Annelie Carlsson, Lotta Högberg, Olof Sandström, Lars Stenhammar, Anneli Ivarsson, and Lars Lindholm. "Is mass screening for coeliac disease a wise use of resources? A health economic evaluation." *BMC gastroenterology* 21, no. 1 (2021): 1-10.
- [12] Black, Christopher J., Douglas A. Drossman, Nicholas J. Talley, Johannah Ruddy, and Alexander C. Ford. "Functional gastrointestinal disorders: advances in understanding and management." *The Lancet* 396, no. 10263 (2020): 1664-1674.
- [13] Youssefi, Masoud, Mohsen Tafaghodi, Hadi Farsiani, Kiarash Ghazvini, and Masoud Keikha. "Helicobacter pylori infection and autoimmune diseases; Is there an association with systemic lupus erythematosus, rheumatoid arthritis, autoimmune atrophy gastritis and autoimmune pancreatitis? A systematic review and meta-analysis study." *Journal of Microbiology, Immunology and Infection* 54, no. 3 (2021): 359-369.
- [14] Francavilla, Ruggiero, Fernanda Cristofori, Mirco Vacca, Michele Barone, and Maria De Angelis. "Advances in understanding the potential therapeutic applications of gut microbiota and probiotic mediated therapies in celiac disease." *Expert Review of Gastroenterology & Hepatology* 14, no. 5 (2020): 323-333.
- [15] Aaron, Lerner, Matthias Torsten, and Wusterhausen Patricia. "Autoimmunity in celiac disease: extra-intestinal manifestations." *Autoimmunity reviews* 18, no. 3 (2019): 241-246.
- [16] Glissen Brown, Jeremy R., and Prashant Singh. "Coeliac disease." *Paediatrics and international child health* 39, no. 1 (2019): 23-31.
- [17] von Mühlenbrock-Pinto, C., and A. M. Madrid-Silva. "Celiac disease in Chilean adults." *Revista de Gastroenterología de México (English Edition)* (2022).
- [18] Kumral, Dennis, and Sana Syed. "Celiac disease screening for high-risk groups: Are we doing it right?." *Digestive Diseases and Sciences* 65, no. 8 (2020): 2187-2195.
- [19] Cichewicz, Allie B., Elizabeth S. Mearns, Aliko Taylor, Talia Boulanger, Michele Gerber, Daniel A. Leffler, Jennifer Drahos, David S. Sanders, Kelly J. Thomas Craig, and Benjamin Lebwohl. "Diagnosis and treatment patterns in celiac disease." *Digestive diseases and sciences* 64, no. 8 (2019): 2095-2106.
- [20] Koh, Joel En Wei, Simona De Michele, Vidya K. Sudarshan, V. Jahmunah, Edward J. Ciaccio, Chui Ping Ooi, Raj Gururajan et al. "Automated interpretation of biopsy images for the detection of celiac disease using a machine learning approach." *Computer Methods and Programs in Biomedicine* 203 (2021): 106010.
- [21] Gupta, S., Kaushal, A., Kumar, A. and Kumar, D., 2017. Ultrasensitive transglutaminase based nanosensor for early detection of celiac disease in human. *International journal of biological macromolecules*, 105, pp.905-911.
- [22] Rosales-Rivera, L.C., Dulay, S., Lozano-Sánchez, P., Katakis, I., Acero-Sánchez, J.L. and O'sullivan, C.K., 2017. Disulfide-modified antigen for detection of celiac disease-associated anti-tissue transglutaminase autoantibodies. *Analytical and bioanalytical chemistry*, 409(15), p.3799.
- [23] Smarrazzo, A., Misak, Z., Costa, S., Mičetić-Turk, D., Abu-Zekry, M., Kansu, A., Abkari, A., Bouziane-Nedjadi, K., Hariz, M.B., Roma, E. and Velmishi, V., 2017. Diagnosis of celiac disease and applicability of ESPGHAN guidelines in Mediterranean countries: a real life prospective study. *BMC gastroenterology*, 17(1), pp.1-8.
- [24] Erlichster, M., Tye-Din, J.A., Varney, M.D., Skafidas, E. and Kwan, P., 2018. Rapid, loop-mediated isothermal amplification detection of celiac disease risk alleles. *The Journal of Molecular Diagnostics*, 20(3), pp.307-315.
- [25] Joelson, A.M., Geller, M.G., Zylberberg, H.M., Green, P.H. and Lebwohl, B., 2019. Numbers and features of patients with a diagnosis of celiac disease without duodenal biopsy, based on a national

survey. *Clinical Gastroenterology and Hepatology*, 17(6), pp.1089-1097.

disease risk prediction: a comprehensive review. *Journal of Food Science and Technology*.

- [26] Tangermann, P., Branchi, F., Itzlinger, A., Aschenbeck, J., Schubert, S., Maul, J., Liceni, T., Schröder, A., Heller, F., Spitz, W. and Möhler, U., 2019. Low sensitivity of simtomax point of care test in detection of celiac disease in a prospective Multicenter Study. *Clinical Gastroenterology and Hepatology*, 17(9), pp.1780-1787.
- [27] Li, B.N.N., Wang, X., Wang, R., Zhou, T., Gao, R., Ciaccio, E.J. and Green, P.H., 2019. Celiac disease detection from videocapsule endoscopy images using strip principal component analysis. *IEEE/ACM transactions on computational biology and bioinformatics*.
- [28] Koh, J.E.W., Hagiwara, Y., Oh, S.L., Tan, J.H., Ciaccio, E.J., Green, P.H., Lewis, S.K. and Acharya, U.R., 2019. Automated diagnosis of celiac disease using DWT and nonlinear features with video capsule endoscopy images. *Future Generation Computer Systems*, 90, pp.86-93.
- [29] Wang, X., Qian, H., Ciaccio, E.J., Lewis, S.K., Bhagat, G., Green, P.H., Xu, S., Huang, L., Gao, R. and Liu, Y., 2020. Celiac disease diagnosis from videocapsule endoscopy images with residual learning and deep feature extraction. *Computer methods and programs in biomedicine*, 187, p.105236.
- [30] Gheshlagh, R.G., Rezaei, H., Goli, M., Ausili, D., Dalvand, S., Ghafouri, H. and Dehkordi, A.H., 2020. Prevalence of celiac disease in Iranian patients with type 1 diabetes: A systematic review and meta-analysis. *Indian Journal of Gastroenterology*, pp.1-7.
- [31] Russo, C., Wolf, R.L., Leichter, H.J., Lee, A.R., Reilly, N.R., Zybert, P., Green, P.H. and Lebowitz, B., 2020. Impact of a Child's Celiac Disease Diagnosis and Management on the Family. *Digestive Diseases & Sciences*, 65(10).
- [32] Ralbovsky, N.M. and Lednev, I.K., 2021. Analysis of individual red blood cells for Celiac disease diagnosis. *Talanta*, 221, p.121642.
- [33] Koh, J.E.W., De Michele, S., Sudarshan, V.K., Jahmunah, V., Ciaccio, E.J., Ooi, C.P., Gururajan, R., Gururajan, R., Oh, S.L., Lewis, S.K. and Green, P.H., 2021. Automated interpretation of biopsy images for the detection of celiac disease using a machine learning approach. *Computer Methods and Programs in Biomedicine*, 203, p.106010.
- [34] Tapkire, M. and Arun, V., 2022. Application of artificial intelligence to correlate food formulationsto