*Grenze International Journal of Engineering and Technology, June Issue*

# Optimized Prediction and Classification of Celiac Disease in Biopsy Images using Transfer Learning

Mayura Tapkire[1] and Vanishri Arun[2]
[1]Department of Information Science and Engineering, The National Institute of Engineering, Mysuru.
[2]Research Scholar at Department of Information Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka, India
Email: vanishriarun@jssstuniv.in

***Abstract*— A biopsy image analysis is necessary for a timely and accurate diagnosis of celiac disease (CD), a chronic autoimmune disease. This research introduces a refined method for predicting and categorizing CD by integrating transfer learning with advanced optimization strategies to improve diagnostic accuracy. The proposed approach follows a structured optimization sequence, beginning with Enhanced Cuttlefish Optimization (ECO) to minimize noise, followed by Improved Whale Optimization (IWO) to extract key texture and shape-related features, and concluding with Modified Crow Search (MCS) for selecting the most relevant features. For classification, these refined characteristics are subsequently fed into a convolutional neural network (CNN) that uses transfer learning. According to experimental results, this approach performs better than traditional methods in terms of accuracy, precision, recall, and F1-score, making it a solid framework for clinical applications involving early CD identification and classification.**

***Keywords*—Celiac disease prediction, biopsy image classification, transfer learning, optimization, convolutional neural network (CNN), ECO algorithm, IWO algorithm, MCS algorithm, feature extraction, medical imaging.**

## I. INTRODUCTION

An unfavorable immune reaction to gluten causes celiac disease (CD), a chronic autoimmune condition that damages and inflames the small intestine. Individuals affected by CD often suffer from a variety of health complications, including neurological, endocrine, reproductive, hematologic, and cardiovascular issues. The disease primarily affects the mucosal lining of the small intestine, particularly the duodenum and jejunum, due to immune-mediated reactions. This process results in the flattening of intestinal villi, which significantly hampers nutrient absorption. Mucosal anomalies include fissures, a mosaic-like pattern, and scalloping of the mucosal folds are frequently discovered during endoscopic examinations.

Tissue transglutaminase (tTG) antibody detection, endoscopic evaluation, and duodenal biopsy are the gold standards for diagnosing CD. However, non-invasive diagnostic methods often fail to detect abnormalities in the distal small intestine, leading to undiagnosed cases. CD is a genetically predisposed condition in which gluten consumption triggers an immune response, resulting in chronic inflammation of the intestinal lining. It is one of the most widespread chronic enteropathies, particularly in Western populations.

*Grenze ID: 01.GIJET.11.2.733*

Studies suggest that a substantial number of cases remain undiagnosed, with many individuals experiencing mild gastrointestinal symptoms or vitamin deficiencies, making diagnosis more complex. The overlap of nonspecific gastrointestinal symptoms with other functional and pathological conditions further complicates clinical identification.

In recent decades, CD has received considerable attention in medical study due to the rising prevalence of autoimmune illnesses such as Crohn's disease, systemic lupus erythematosus (SLE), and rheumatoid arthritis (RA). In people who are genetically susceptible, eating gluten causes the condition by causing autoantibodies to be produced, which harm the intestinal mucosa. Serological markers including anti-gliadin antibodies (AGA) and anti-tTG antibodies are frequently used in conjunction with biopsy, which is still the gold standard for diagnosis. Immunoglobulin A (IgA) is considered the most specific marker; however, a small percentage of patients (2–5%) exhibit selective IgA deficiency, necessitating the measurement of IgG subclass antibodies for accurate detection. Symptoms of CD vary widely, ranging from common issues like diarrhea and weight loss to less typical indicators such as elevated liver enzyme levels and chronic fatigue. Current diagnostic approaches, which rely on symptom-based case identification, often result in misdiagnoses, particularly in atypical presentations. This highlights the urgent need for improved diagnostic strategies to address the high prevalence of undiagnosed symptomatic CD.

### *A. Research Contributions*

To tackle these challenges, we present a hybrid machine learning framework (HML-PCD) for the early and precise prediction of celiac disease (CD). The primary objectives of HML-PCD are:

- Preprocessing Enhancement – Implementing an optimization-driven preprocessing technique to refine biopsy images and improve detection accuracy.
- Feature Extraction – Developing an advanced optimization method to capture diverse feature types, including texture and non-linear attributes, for better disease prediction.
- Feature Selection – Proposing an efficient feature selection approach to minimize data complexity while retaining essential information for classification.
- Transfer Learning – Designing a model capable of both binary and multi-class classification to strengthen disease detection and categorization.

The sections of this work that follow are organized as follows: Related work on CD diagnosis and classification is reviewed in Section 2. The problem formulation and system design are explained in Section 3. The suggested approach is explained in Section 4. The experimental results and comparison analysis are presented in Section 5, and the study's conclusion and possible future research areas are covered in Section 6.

## II. RELATED WORKS

The diagnosis and categorization of celiac disease (CD) have been the subject of several investigations in recent years. Table 1 summarizes the methodologies, parameters, and future scopes discussed in the literature.

One study utilized a hybrid comprising AuNP-enhanced GQD/PAMAM for CD detection [21]. The approach employed EDC-NHS cross-linking chemistry to bond AuNP encapsulated within MWCNT with GQDs, leveraging cysteine thiol and amino groups. Similarly, Rosales et al. [22] introduced an immune sensor to identify anti-tissue transglutaminase antibodies. The method involved chemically attaching disulfide groups to antigenic proteins for IgA and IgG autoantibody detection. Both enzyme-linked immunosorbent assays and surface plasmon resonance (SPR) techniques were applied to analyze protein antigenicity.

Smarrazzo et al. [23] examined how children from 13 Mediterranean countries were diagnosed with CD using the ESPGHAN standards. Patients at 14 locations had thorough diagnostic work-ups as part of this study. To determine the genotypes of HLA-DQ2.5, HLA-DQ8, HLA-DQ2.2, and HLA-DQA1*05, Erlichster et al. [24] created CD-LAMP. This technique showed 100% accuracy in 40 blood and 20 saliva samples without the requirement for heating or DNA purification, and it had great sensitivity (97.1%) and specificity for the DQ8 and DQ2.2 genotypes.

Patient-powered research network (PPRN) was recommended for comparing biopsy-diagnosed CD cases with serology-diagnosed cases [25]. A computer-aided diagnosis (CAD) system that extracted nonlinear and textural information from endoscopic images using the discrete wavelet transform (DWT) was proposed by Koh et al. [28].

Wang et al. [29] introduced a unique deep learning calibration module featuring a block-wise recalibration component to enhance local channel feature map analysis. Their model integrated VCE image analysis with ResNet50 and Inception-v3 to separate CD images from healthy controls using classifiers such as LDA, k-NN,

and SVA. Gheshlagh et al. [30] conducted a meta-analysis to determine the prevalence of CD in Iranian individuals with type 1 diabetes using Cochran's Q test for heterogeneity analysis.
Russo et al. [31] used quality-of-life evaluations and semi-structured interviews to develop a novel diagnostic method for CD in children. Furthermore, Ralbovsky et al. investigated a low-cost, non-invasive diagnostic technique that compares the shape of red blood cells from CD donors with gluten-free controls using Raman hyperspectroscopy. Their study employed partially least squares discriminating analysis (PLS-DA) for analyzing Raman spectra data.
The research on celiac disease (CD) has highlighted several methodologies and their respective gaps. In [21], the use of MWCNT with EDC-NHS cross-linking chemistry demonstrated stable sensor electrode performance at 4°C, allowing for illness detection without requiring a new electrode. The study in [22] employed disulfides and rabbit polyclonal antibodies, recommending the development of electrochemical immune sensors for real-world patient sample detection. The ESPGHAN guidelines applied in [23] suggested that CD diagnosis could be considered without requiring an intestinal biopsy, emphasizing family discussions. Meanwhile, CD-LAMP in [24] provided a highly accurate, rapid, and cost-effective clinical assessment when paired with serology.
A patient-powered research network (PPRN) was proposed in [25] to enhance evidence-based diagnosis and treatment. For point-of-care testing (POCT), the Simtomax method in [26] identified the need for further improvement. The SPCA with a non-greedy L1-norm and a non-parametric k-NN method in [27] was suggested to evaluate the performance of gluten-free diets in CD patients. A CAD system in [28] leveraged DWT to improve diagnostic and treatment timing for CD patients. The deep learning recalibration module proposed in [29] emphasized exploring newer algorithms and larger datasets for faster and more efficient diagnosis and therapy.
The study in [30] highlighted the need for appropriate screening procedures to mitigate the risks of osteoporosis and cancer in CD patients. Diagnosis methods for children in [31] underlined the importance of research in broader family demographics, particularly among low-income families and regions. Finally, the diagnostic method using PLS-DA in [32] was recommended for validation with larger sample sizes in future studies.
WeiKoh et al. [33] employed modified Marsh ratings to diagnose and categorise celiac disease using machine learning. Sub bands were created using the SPT method, from which entropy and nonlinear properties were calculated.

## III. Methodology and System Architecture

### *A. Background*

For those who are genetically susceptible, gluten ingestion can cause celiac disease, an autoimmune disorder. When the small intestine is damaged by this immune response, symptoms can range from mild digestive pain to more serious issues like anemia, tiredness, and neurological abnormalities. Long-term health problems, such as malnourishment, osteoporosis, difficulties conceiving, and an increased risk of several types of cancer, can result from untreated celiac disease. Although diagnostic methods like serological tests and biopsies are available, accurately identifying the disease in its early stages remains difficult due to its varied clinical symptoms and preclinical progression. Developing non-invasive, cost-effective diagnostic tools for early detection is therefore essential.
Machine learning techniques provide a promising approach by analyzing extensive datasets to uncover patterns that conventional diagnostic methods may overlook. Hybrid machine learning models, which integrate multiple algorithms, can enhance diagnostic accuracy by incorporating genetic markers, clinical symptoms, and test results. This approach enables earlier detection and more precise classification of the disease. This chapter examines how hybrid machine learning techniques can contribute to improving early diagnosis and classification of celiac disease, with a particular focus on deep learning models to address its complexity and refine diagnostic strategies.
This revision makes the text more natural and original while maintaining its technical

### *B. Observations*

Early diagnosis is essential to preventing serious complications like malnutrition, osteoporosis, and gastrointestinal cancers in cases of celiac disease, an autoimmune condition brought on by gluten consumption. Current diagnostic methods like serological assays and biopsies are often invasive, time-consuming, and ineffective at identifying early or atypical cases. This study aims to develop a non-invasive, accurate system for early celiac disease prediction and classification by applying a combination of supervised and unsupervised learning, along with feature selection and ensemble techniques. The system will utilize patient data, including

demographics, genetics, serological results, and clinical features, to provide alerts and assess the risk of celiac disease. The key objectives are to identify important predictive factors, develop a hybrid machine learning model, and evaluate its performance using metrics like F1-score, precision, recall, and accuracy.

*C. System design and Dataset Structure*

In order to forecast and categorize celiac disease, the suggested system will work as a pipeline, including different parts for data collecting, pre-processing, feature extraction, model construction, and deployment. The high-level system design includes the following stages:

- Data Collection: Collecting biopsy images to form the dataset.
- Data Pre-processing: Cleaning the data by removing missing, inconsistent, and outlier values, using the Enhanced Cuttlefish Optimization (ECO) algorithm, which is based on the Cuttlefish Optimization algorithm.
- Feature Extraction: Extracting texture and non-linear features using the Improved Whale Optimization (IWO) method to identify disease-related characteristics.
- Feature Selection: Applying the Modified Crow Search (MCS) algorithm to select the most relevant features and reduce dimensionality.
- Prediction and Classification: Inputting new patient data into the system, which outputs a classification of celiac disease risk using a transfer learning approach with ResNet-50 [38] to enhance detection accuracy. A flowchart illustrating the system design is provided in Figure 1.

High-quality image datasets are crucial for developing automated systems that support accurate classification of conditions using machine learning (ML) and deep learning (DL) methods. A dataset containing representative images of celiac disease, non-celiac disease (control or normal cases), and nonspecific duodenitis is essential for training and building ML models aimed at early diagnosis and classification.

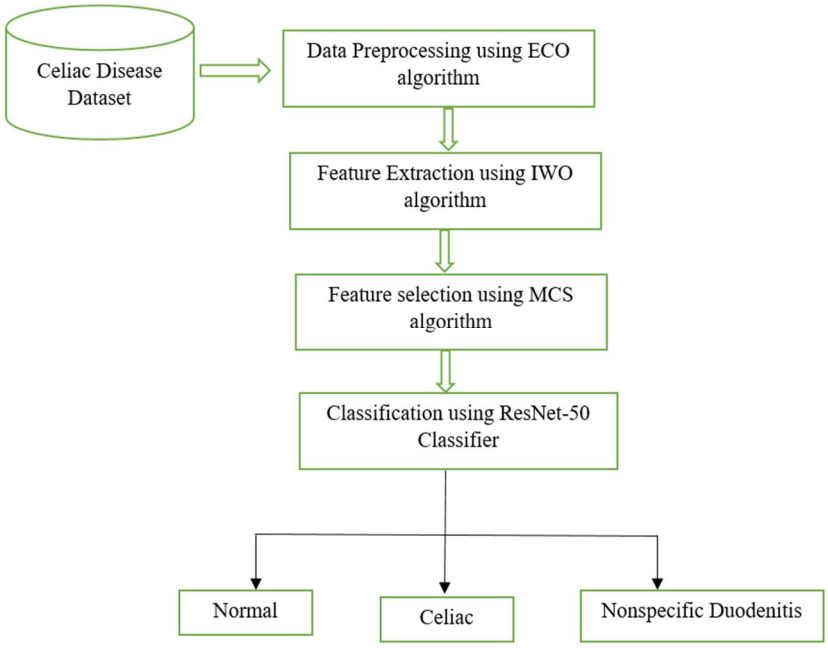


Figure 1. System design of proposed technique

*D. Proposed Methodology*

The purposes of feature extraction, optimal feature selection, and data pre-processing are explained in the section on the suggested technique. Lastly, we provide definitions for illness forecasting and classification.

*Utilizing the Enhanced Cuttlefish Optimization (ECO) Algorithm for Data Preprocessing*

The method of Cuttle Fish Optimization (CFO) resolves numerical global optimization problems by mimicking cuttlefish's ability of color changing. The colors and patterns of a cuttlefish are produced by the reflection of light through three cell layers. Evolutionary algorithms of artificial intelligence systems are focused on genetic evolution. CFO is a bioinspired optimization technique that mimics the color-changing ability of cuttlefish. The Enhanced Cuttlefish Optimization (ECO) method, which is one of the improvements of the traditional approach.

In the beginning, the two parameters, $R_1$ and $R_2$ are used to construct the reflection strategy. For visibility, the two parameters are denoted as $U_1$ and $U_2$. Further, the random population denoted as $Q$ is represented as-

$$Q[y_i].\ point\ s[i] = random\ (Upper\ limit - Lower\ limit) + Lower\ limit \quad (3.1)$$

A vector of continuous values signifies a certain value. In the case of this study, a fitness value and associated coordinates are part of this vector, and $y_j$ displays the best optimal solution obtained after the current iteration.

$$H_{Size} = Q/4 \quad (3.2)$$

In which $H_{Size}$ represents the size of each group and the initial population size is denoted as Q. During the decision-making stage, equation 3.3 is employed in generating new results several times until the last requirements are fulfilled.

$$y_{new} = reflection + visibility \quad (3.3)$$

in which $y_{new}$ is the new solution that is denoted as sum of reflection and visibility. Before the evaluation of the $y_{new}$, the Reflection $R$ and $U$ needs to be determined. In the above equation that follows, the $R$ is fixed to 1. In the equation 3.1 and 3.2 the value of $U$ is 1.

$$r = random\ () * (R_1 - R_2) + R_2 \quad (3.4)$$

$$u = random\ () * (U_1 - U_2) + U_2 \quad (3.5)$$

where *random()* denotes a random number from 0 to 1. Further, R1, R2, U1 and U2 are set initially to any value. To obtain a best score, the $Bu_{best}$ is calculated as

$$Bu_{best} = \frac{\sum[bestsolution.poibts[i]]}{N} \quad (3.6)$$

in which the $Bu_{best}$ is the best solution and is calculated by dividing by N, where N is the total score which reflects sum of the calculated reflections visibility.

Firstly, experiments are conducted using the four search algorithms on the subsets of data listed below. Applying equations (3.7) and (3.8), a new solution $x_{new}$ is calculated. This is a global search and value of $r$ is obtained using equation (3.4):

$$reflection = r * H_1[y_i].point[i] \quad (3.7)$$

$$visibilty = u * (bestsolution.point[i] - H_1[y_j].point[i]) \quad (3.8)$$

where $H_1$ is the first set of solutions, $y_j$ is the second solution in the group. $[y_i].point\ s[i]$ shows point $i$ of the solution. A solution was developed as described in (3.9) and (3.10) for the reflection and visibility. Further, $H_2$ acts as a local search and calculate optimal solution using the value $u$.

$$reflection = r * bestsolution.point[i] \quad (3.9)$$

*Algorithm 1 Data pre-processing using ECO algorithm*

The pseudocode for the Enhanced Cuttlefish Optimization (ECO) method is illustrated below.

***Input***: $Q$ points, $y_j$

***Output***: $H_{Size}$ and $Bu_{best}$

*Initialize the population Q*

*Initialize parameters R1, R2, U1 and U2*

*Calculate fitness* $H_{Size} = Q/4$

*Calculate best solution*

***if*** *i = 0 and j = 1*

*Calculate y values as*

*Calculate* $y_{new}$ *= reflection + visibility*

*Calculate* ***r*** *as random () * ($R_1 - R_2$) + $R_2$*

***u*** *as random () * ($U_1 - U_2$) + $U_2$*

*Update position using activation functions*

***end if***

***End***

A updated new solution is obtained as $y_{new}$. The equation (3.9) and (3.10) is used for calculation of new values of *visibility* and *reflection*.

$$visibilty = u * (bestsolution.po\,\text{int}\, s[i] - Bu_{best} \tag{3.10}$$

For instance, If there are two points in the best solution (6, -2), U1 and U2 are set to (1,-1), respectively. The result of the $Bu_{best}$ will be calculated as $Bu_{best}$ = (6 - 2) / 2 = 2.
Also, the $H_{5\text{-}Size}$ is calculated as-

$$H_{5-size} = \left[ \left( (H5T_{percent} * H_{size}) * 4 \right) + \frac{H_{size}}{2} \right] \tag{3.11}$$

Only the best options are chosen, and the outcomes differ from the first four groups.
$H_{5\text{-}Size}$ as follows:

$$H_5 = (\uparrow H_1 * H5T_{percent}) + (\uparrow H_2 * H5T_{percent}) + (\uparrow H_3 * H5T_{percent}) + (\uparrow H_4 * H5T_{percent}) \tag{3.12}$$

where ↑ represents the sorted solutions from best to worst. Algorithm 1 describes the working function of Enhanced Cuttlefish Optimization (ECO) algorithm.

*The Improved Whale Optimization (IWO) Algorithm for Multi-level Feature Extraction*

Feature extraction is an indispensable approach applied in computer visions in the case of many advanced applications. Gathering every potential feature for a given dataset (training set) is part of the multi-level feature extraction process of the Improved Whale Optimization method (IWO). The multi-level feature extraction aims to achieve these objectives in order to reduce the complexity of training the classification strategy implemented. The Whale Optimization method is a kind of optimization strategy that mimics humpback whale hunting activities and is categorized under meta-heuristic techniques. The bubble-net hunting technique served as the model for the algorithm. Humpback whales adopt a foraging technique known as bubble-net feeding.
In the IWO algorithm, two real numbers are generated $R_1$ and $R_2$. Further the values *B* and *D* are calculated using the formula as shown below.

$$\vec{B} = 2 \cdot \vec{b} \cdot \overrightarrow{R_1} - \vec{b} \tag{3.13}$$

$$\vec{D} = 2 \cdot \overrightarrow{R_2} \tag{3.14}$$

where *b* indicates a real number from 0 to 2 that is considered in every iteration. In this strategy, the search agent is obliged to vacate its current location and move about the environment in search of a target. The position update function is therefore defined as follows.

$$\vec{C} = \left| \vec{D} \times \overrightarrow{Q_{rand}} - \overrightarrow{Q_L^s} \right| \tag{3.15}$$

$$\overrightarrow{Q_L^{s+1}} = \overrightarrow{Q_{rand}} - \vec{B} \times \vec{C} \tag{3.16}$$

where $Q_{rand}$ is approximately the generated level vector on the boundary line. Also, the notation $Q_L^{s+1}$ in the *i-th* slot is the generation vector of the search agent status vector and $Q_L^{s+1}$ is the process entails the evolution of the *S + 1* vector of the search agent status vector. Throughout this process, a position of a target is established, and the target is surrounded. The agent doing the search is nearly in the right position. The following is a capture and illustration of the position update.

$$\vec{C} = \left| \vec{D} \times \overrightarrow{Q^s} - \overrightarrow{Q_L^s} \right| \tag{3.17}$$

$$\overrightarrow{Q_L^{s+1}} = \left| \overrightarrow{Q_L^s} - \vec{B} \times \vec{C} \right| \tag{3.18}$$

where $Q_*^S$ is the position vector of optimal search agent's generation *t*. In this case, the search agent is forced to vacate its current position and to move randomly about in space in search of food using this strategy. The position transformation formula and mathematical model is described as follows.

$$\overrightarrow{Q_L^{s+1}} = \overrightarrow{C^s} \cdot E^{aL} \cdot \cos(2\pi L) + \overrightarrow{Q_*^s} \tag{3.19}$$

In the case of a logarithmic helix, the logarithmic shape constant is any real number in the interval [-1,1]. Regarding the random stabilization phase, the cochlear mutation is suppressed by using the search agent's classical Cauchy's functionality. When used, the general function's defining expression looks like this:

$$f^{-1}(q; y_0, \gamma) = y_0 + \gamma \cdot \tan(\pi \cdot (q - 1/2)) \tag{3.20}$$

Therefore, the following is an update to the Whale Optimization technique for random prey:

$$\overrightarrow{Q_L^{s+1}} = \overrightarrow{q_L^s} + \vec{B} \cdot \tan(\pi \cdot (R_3 - 1/2)) \tag{3.21}$$

where j denotes the index of the position of search agents $Q_L^S$ in generation of *S*, which is a uniformly distributed random value in the range of [0, 1]. A local mutation probability *K* is provided to enhance the stability of the algorithm. When the above operation is performed, the fittest surviving global-best search agent is retained, and the following equation is obtained as:

$$\overrightarrow{Q_{*M}^s} = \begin{cases} \overrightarrow{Q_*^s} \cdot (1+n) & , R_4 > K \\ \overrightarrow{Q_*^s} & ,, R_4 \le K \end{cases} \tag{3.22}$$

The algorithm for the IWO is described in the below section.

*Algorithm 2 for Improved Whale Optimization*

The pseudocode for the Improved Whale Optimization (IWO) method is illustrated below.

***Input***: *B and D*

***Output***: $Q_*^S$

*Generate the initial population*

*Evaluate the fitness for each candidate*

***while*** *the halting criterion is not satisfied* ***do***

***for*** *j = 1 to* $n_q$

*Update the values* $\overrightarrow{Q_L^{s+1}} = \overrightarrow{Q_{rand}} - \vec{B} \times \vec{C}$

*Update the boundary range as* $\overrightarrow{Q_L^{s+1}} = \overrightarrow{C^s} \cdot E^{aL} \cdot \cos(2\pi L) + \overrightarrow{Q_*^s}$

*Use the Whale Optimization as* $\overrightarrow{Q_L^{s+1}} = \overrightarrow{q_L^s} + \vec{B} \cdot \tan(\pi \cdot (R_3 - 1/2))$

*Use the global search for better fitness as* $\overrightarrow{Q_{*M}^s} = \begin{cases} \overrightarrow{Q_*^s} \cdot (1+n) & , R_4 > K \\ \overrightarrow{Q_*^s} & ,, R_4 \le K \end{cases}$

*end* ***for***

*end* ***while***

*The Modified Crow Search (MCS) Algorithm for Optimal Feature Selection*

A feature selection method is developed in optimal way after feature extraction in order to rank features in the dataset and identify the best of them, reducing the number of false negative features accepted to half and keeping the same level true positive rates. This means that it is more effective in identifying the right variables making the model less tedious and complicated and more precise. The Crow Search (CS) technique is a form of evolutionary computation based on population that imitates the actions and interactions of crows with one another. Crows are birds that are highly intelligent within their body sizes, exhibit social behaviors, have large

enough brains, and are known to stash food in places in their surroundings for days, weeks or even months. The modified Crow Search method was conceived by the use of chaos theory on the original algorithm for the Crow Search Method (CSM). The CSM approach does not require the crow i to be pursued, but finds and changes its position in respect to the reservoir of food of crow j. The same is represented as below.

$$p_j^{(s+1)} = p_j^{(s)} + R_j * fL_i^{(s)} * (M_i^{(s)} - p_j^{(s)}), \tag{3.23}$$

where the $f_L$ indicates the length of the flight and $R_j$ indicates a random number between [0,1]. When crow *i* becomes aware of this, the crow *j* attempts to find its food. In this situation, the crow *i* moves erratically to confuse the crow *j*. Both scenarios may be represented mathematically by combining the two in the following manner:

$$p_j^{(s+1)} = \begin{cases} p_j^{(s)} + R_j * fL_i^{(s)} * (M_i^{(s)} - p_j^{(s)}), & R_i \geq Bq_j^s \\ choose\, b\, random\ position, & otherwise \end{cases} \tag{3.24}$$

During the course of the algorithm, each crow is evaluated using a specific fitness function. The crows then relocate depending on the fitness rate calculated for each individual crow. Each new task is assessed for its potential effectiveness. Updated information is integrated into the crows' memories as:

$$M_j^{(s+1)} = \begin{cases} p_j^{(s)} \ \ if\ f(M_j^{(s+1)})\ is\ better\ than\ f(p_j^{(s)}) \\ M_j^{(s+1)}, \ \ otherwise \end{cases} \tag{3.25}$$

The collection of values is expressed using two dimensions, in which a solution may be referred to in terms of a binary vector description of values 0 and 1. This entails switching the agents from the fixed space to the binary one.

$$p_j^{(s+1)} = \begin{cases} 1 \ \ if\ \ (t(p_j^{(s+1)})) \geq rand() \\ 0, \ \ otherwise \end{cases} \tag{3.26}$$

$$where\, t = \frac{1}{1 + E^{10}(p_j^{(s+1)} - 0.5)}$$

Any perturbation in the system may induce subsequent actions in a non-linear manner. Use of random variables is found in CSA to alter the positioning of the crow. In the chaos phenomenon, any perturbation in the system may cause non-linear variations in actions that follow. The random factors utilized in order to change the crow position in the algorithm are listed below.

$$p_j^{(s+1)} = \begin{cases} p_j^{(s)} + D_i * (M_i^{(s)} - p_j^{(s)}), & D_w \geq Bq_j^s \\ choose\, b\, random\ position, & otherwise \end{cases} \tag{3.27}$$

where $D_i$ and $D_w$ indicates the value obtained from the chaotic map after the *i*-th iteration and *w*-th iteration respectively. Further, the $U_{Shape}$ is obtained as in (3.27).

$$U_{shape} = \left| \frac{2}{\pi} \arctan(\frac{\pi}{2} p_j^s) \right| \tag{3.28}$$

Consequently, it has become possible to formulate the fitness function in such a way that solutions with respect to the two objectives are targeted.

$$Fitness = \alpha \Delta_r(C) + \beta \frac{|X|}{|S|} \tag{3.29}$$

The algorithm for the Modified Crow Search algorithm is described as.

*Algorithm 3 Optimal feature Selection Using MCS Algorithm*

***Input*** *Extracted features*

***Output*** *Optimally selected best features*

*Initialize crows' population* $R_j$

*Compute the fitness function as*

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

*Update the CSA as* $p_j^{(s+1)} = \begin{cases} p_j^{(s)} + R_j * fL_i^{(s)} * (M_i^{(s)} - p_j^{(s)}), & R_i \geq Bq_j^s \quad with \\ choose\, b\, random\ position, & otherwise \end{cases}$

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

*boundary range defined as*

*Check the feasibility of the* $M_j^{(s+1)} = \begin{cases} p_j^{(s)} & if\ f(M_{shape}^{(s+1)}) is\ better\ than\ f(p_j^{(s)}) \\ M_j^{(s+1)}, & otherwise \end{cases}$ *crow as*

*Obtain the best solution as*

***End***

*Disease Prediction and Classification Using Transfer Learning-based ResNet-50*

Over time, physicians have relied on statistical methods, intuition, and clinical experience to assess the likelihood of a disease or estimate patient outcomes.

However, this approach often results in biases, errors, and high costs, negatively affecting the quality of treatment.

With advancements in technology, particularly through the availability of electronic health records, machine learning (ML) algorithms have been applied to healthcare, enabling more efficient research and clinical decisions.

In this context, Transfer Learning-based Convolutional Neural Networks (CNNs), specifically ResNet-50, represent a significant advancement over traditional models.

ResNet-50, which uses residual blocks to ease the training of deep networks, is particularly useful for medical image classification tasks, as it helps capture hierarchical features in images, which is essential for accurate diagnosis.

The ResNet-50 model is fine-tuned using domain-specific medical data (e.g., biopsy images for Celiac Disease) after being pre-trained on a sizable dataset like ImageNet.

This approach reduces the need for a large amount of labelled data and accelerates the model's learning process, as the pre-trained model already possesses rich feature representations from general images.

The transfer learning method enhances the model's capacity to generalize to new data and reduces overfitting.

ResNet-50 CNN Algorithm Steps:

- Input: Preprocessed image data with optimal features.
- Load Pre-trained ResNet-50 model.
- Fine-tune the ResNet-50 model:
  - Replace the final layer with a suitable classifier layer.
- Train the model using the training set:
  - Use of data augmentation and normalizing techniques.
- Calculate the loss function (e.g., Cross-Entropy).
- Update weights via backpropagation using an optimizer.
- Evaluate model performance on validation data.
- Predict CD risk on test data using the trained model.

## IV. Results Discussion

We used a well-known dataset to evaluate our transfer learning approach's performance. The accuracy, precision, recall, F1-score, G-mean, and AUC of the proposed transfer learning strategy were compared to the then-current state-of-the-art methods. Three popular Python 3.7 environments—Anaconda Navigator, Spyder, and Jupyter—were employed for the experiment.

### *A. Dataset*

Biopsy slides from both CD patients (34 slides) and healthy people (63 slides) were available from the University of Virginia (UVa) archives. The demographics of the study participants were as follows: approximately 52% (n = 53) were boys and girls, and 48% (n = 49) were adults, with an average age of 37.5 months (ranging from 19.0 to 121.5 months). CD patients represented 37.7% of the participants (51.8%). From 63 CD patients, UVa received 239 duodenal biopsy slides stained with H&E. The dataset includes 7,125 patches for Type 1 illness, 6,842 patches for Type 2, 8,120 patches for Type 3, and 8,111 patches for Type 4. The study used two sets of training and testing data for both parent and child models. For the parent model, the training set includes 21,140 patches, with 9,058 testing image patches. The child model dataset contains four types of CD: Type 1 (4,988 training patches and 2,137 test patches), Type 2 (4,790 training patches and 2,052 test patches), Type 3 (5,684 training patches and 2,436 test patches), and Type 4 (5,678 training patches and 2,137 test patches). Table 2 provides further details of the dataset.

TABLE I: DATASET DESCRIPTION

| Celiac disease severity | Number of patient records used | | |
|---|---|---|---|
| | Training | Testing | Total |
| Type 1 | 4988 | 2137 | 7125 |
| Type 2 | 4790 | 2052 | 6842 |
| Type 3 | 5684 | 2436 | 8120 |
| Type 4 | 5670 | 2433 | 8111 |
| The total quantity of patient data | **21140** | **9058** | **30198** |

### *B. Performance Metrics*

In simulation findings, our suggested transfer learning approach performs better than the current MWCNT [21], ESPGHAN [23], CD-LAMP [24], and PLS-DA approaches in terms of accuracy, precision, recall, G-mean, and AUC.

TABLE II: PERFORMANCE COMPARISON

| **Models** | **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 | 81.00 | 95.07 | 94.14 |
| ANN [33] | 96.78 | 95.98 | 94.39 | 98.78 | 93.04 | 95.76 |
| ResNet-50 CNN | 97.85 | 96.78 | 98.12 | 99.14 | 96.78 | 96.08 |

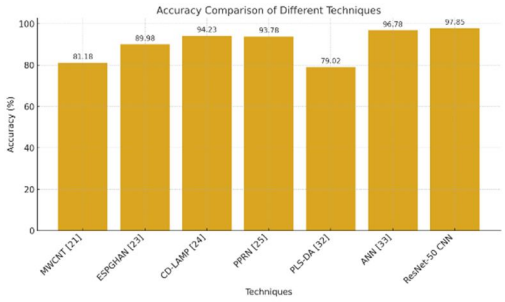


Figure 2 Comparison of proposed and existing models' accuracy

### *C. Comparative analysis*

The detection and classification of illicit activities will be enhanced through the use of the proposed transfer learning approach, as mentioned earlier. Table 3 shows a comparison between the suggested model and current methods. MWCNT [21], ESPGHAN [23], CD-LAMP [24], PPRN [25], PLS-DA [32], and ANN [33] are the current models that are outperformed by 17.036%, 8.043 percentage points, 3.700 percentage points, 4.159 percentage points, 19.244 percentage points, and 1.094 percentage points, respectively, by the ResNet-50 CNN-based model.

The suggested ResNet-50 CNN technique outperforms MWCNT [21], ESPGHAN [23], CD-LAMP [24], PPRN [25], PLS-DA [32], and ANN [33] by 28.679%, 4.637%, 2.375%, 22.503%, and 3.801%, respectively, to achieve a 19.415% improvement in recall. With an F1-score that is 17.279%, 20.405%, 1.785%, 0.968%, 18.297%, and 0.363% higher, respectively, the ResNet-50 CNN model outperforms ANN [33], ESPGHAN [23], CD-LAMP [24], PPRN [25], PLS-DA [32], and MWCNT [21].

## V. CONCLUSION

Globally, the prevalence of celiac disease is rising quickly, requiring medical and technological solutions [37]. In this work, we offer a convolutional neural network (CNN) strategy for automatic early detection of celiac disease based on transfer learning. The following are this work's primary contributions:

Proposed approach utilizes a ResNet-50 CNN model for data preprocessing, effectively removing unwanted artifacts from patient sample data. By employing a transfer learning-based strategy, the model extracts multi-class features, including texture and non-linear feature vectors, which greatly enhance the efficiency of disease detection. Additionally, an optimal feature selection process is incorporated to identify the best features and reduce dimensionality, further improving the model's performance. Ultimately, the ResNet-50 CNN model is applied for celiac disease prediction and classification, significantly boosting detection accuracy.

According on simulation data, the suggested ResNet-50 CNN-based method outperforms current state-of-the-art methods by 12% on average. The suggested model performs 13.45% better overall than the existing techniques. Additionally, compared to current techniques, the ResNet-50 CNN model's F1-score is 17.56% higher. Important performance indicators like AUC, F1-score, and G-mean show how much better the suggested strategy is than current best practices.

Future work can focus on integrating advanced techniques like attention mechanisms or generative models to improve detection accuracy. Testing the model on larger, diverse datasets and incorporating multi-modal data, such as genetic and clinical information, could enhance performance. Additionally, real-world deployment and integration with electronic health records (EHR) could provide valuable feedback for continuous improvement.

## 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-Menabawey, 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., Aliki 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, Aliki 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.
[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 Lebwohl, 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] Eesa, A.S., Brifcani, A.M.A. and Orman, Z., 2014. A new tool for global optimization problems-cuttlefish algorithm. International Journal of Mathematical, Computational, Natural and Physical Engineering, 8(9), pp.1208-1211.
[35] Mirjalili, S. and Lewis, A., 2016. The whale optimization algorithm. Advances in engineering software, 95, pp.51-67.
[36] Askarzadeh, A., 2016. A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm. Computers & Structures, 169, pp.1-12.
[37] Tapkire, M. and Arun, V., 2022. Application of artificial intelligence to corelate food formulationsto disease risk prediction: a comprehensive review. Journal of Food Science and Technology.
[38] V. Sharma and N. Singh, "Deep Convolutional Neural Network with ResNet-50 Learning algorithm for Copy-Move Forgery Detection," 2021 7th International Conference on Signal Processing and Communication (ICSC), Noida, India, 2021, pp. 146-150, doi: 10.1109/ICSC53193.2021.9673422.
[39] Tapkire, M., Arun, V., Lavanya, M.S. et al. Gluten identification from food images using advanced deep learning and transfer learning methods. J Food Sci Technol (2024). https://doi.org/10.1007/s13197-024-06158-y