

# SCIENTIFIC REPORTS



OPEN

## Automatic classification of IgA endomysial antibody test for celiac disease: a new method deploying machine learning

Florentino Luciano Caetano dos Santos<sup>1</sup>, Irmina Maria Michalek<sup>2</sup>, Kaija Laurila<sup>1</sup>, Katri Kaukinen<sup>1,3</sup>, Jari Hyttinen<sup>4</sup> & Katri Lindfors<sup>1</sup>

Widespread use of endomysial autoantibody (EmA) test in diagnostics of celiac disease is limited due to its subjectivity and its requirement of an expert evaluator. The study aimed to determine whether machine learning can be applied to create a new observer-independent method of automatic assessment and classification of the EmA test for celiac disease. The study material comprised of 2597 high-quality IgA-class EmA images collected in 2017–2018. According to standard procedure, highly-experienced professional classified samples into the following four classes: I - positive, II - negative, III - IgA deficient, and IV - equivocal. Machine learning was deployed to create a classification model. The sensitivity and specificity of the model were 82.84% and 99.40%, respectively. The accuracy was 96.80%. The classification error was 3.20%. The area under the curve was 99.67%, 99.61%, 100%, and 99.89%, for I, II, III, and IV class, respectively. The mean assessment time per image was 16.11 seconds. This is the first study deploying machine learning for the automatic classification of IgA-class EmA test for celiac disease. The results indicate that using machine learning enables quick and precise EmA test analysis that can be further developed to simplify EmA analysis.

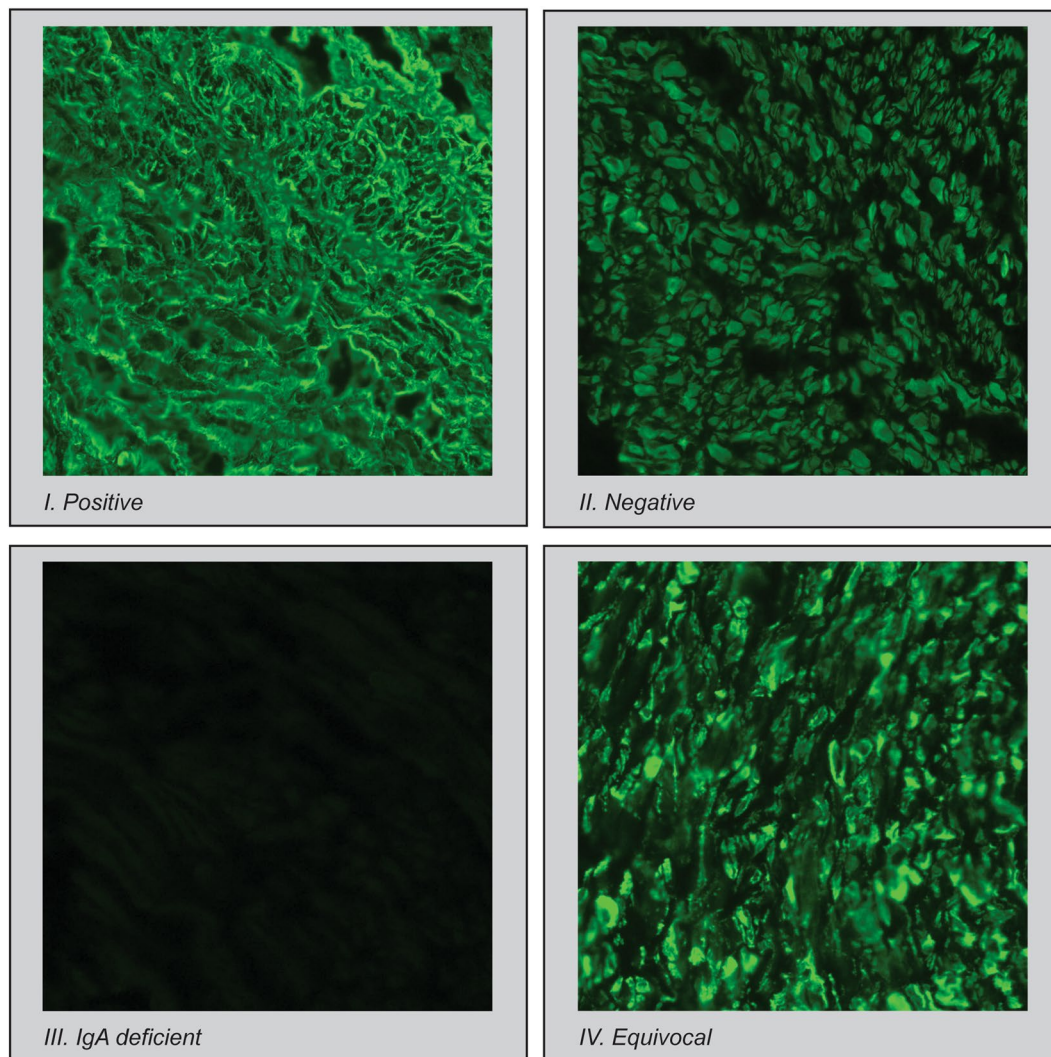
Celiac disease is an immune-mediated enteropathy driven by ingestion of gluten-containing cereals. The clinical course of the disorder is highly variable, and the gluten-induced symptoms may be both gastrointestinal as well as extra-intestinal. The spectrum of the above symptoms can range from mild to severe. Some patients may remain asymptomatic<sup>1</sup>. Such heterogeneity of clinical presentation makes diagnosing celiac disease challenging, and therefore, the disorder remains heavily unrecognized and underdiagnosed worldwide<sup>2</sup>.

Hitherto demonstration of small bowel mucosal villous atrophy, intraepithelial lymphocytosis, and crypt hyperplasia in biopsies obtained upon esophagogastroduodenoscopy has been the cornerstone for the diagnosis of celiac disease. Presence of celiac disease-specific IgA-class autoantibodies, determined by endomysial (EmA) and transglutaminase 2 autoantibody (TG2-Ab) assays, supports the diagnosis and serves as a valuable tool in selecting patients for endoscopy<sup>3,4</sup>. Presence of TG2-Ab in blood samples is detected using immunoassays enzyme-linked immunosorbent assay (ELISA) while the EmA test is an immunofluorescence-based method. The EmA test is regarded as the gold standard when determining the celiac disease autoantibodies, but due to its labor-intensiveness, high cost, and subjective interpretation, its use is limited in clinical practice<sup>3</sup>.

The value of celiac disease-specific autoantibody tests in the diagnostics is systematically acknowledged due to several limitations of the biopsy-based diagnosis, for instance, patchiness of small bowel mucosal lesions, difficulties regarding sampling, processing, and pathomorphological interpretation<sup>5–15</sup>. Moreover, the current celiac disease-specific antibody tests are highly accurate. Both EmA and TG2-Ab tests have excellent sensitivity (90–100%) and specificity (close to 100%)<sup>1</sup>. According to the current European Society for Pediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) guidelines, the diagnosis of celiac disease can be made without a biopsy in symptomatic children with TG2-Abs exceeding 10 times the upper limit of normal in two independent

<sup>1</sup>Celiac Disease Research Center, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.

<sup>2</sup>Faculty of Social Sciences, Tampere University, Tampere, Finland. <sup>3</sup>Department of Internal Medicine, Tampere University Hospital, Tampere, Finland. <sup>4</sup>Computational Biophysics and Imaging Group, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. Correspondence and requests for materials should be addressed to F.L.C.d.S. (email: [florentino.l.c.santos@ieee.org](mailto:florentino.l.c.santos@ieee.org))



**Figure 1.** IgA-class EmA test classes.

| Class |               | Number of samples | Percent of the total dataset |
|-------|---------------|-------------------|------------------------------|
| I     | Positive      | 274               | 10.55%                       |
| II    | Negative      | 2260              | 87.02%                       |
| III   | IgA deficient | 13                | 0.50%                        |
| IV    | Equivocal     | 50                | 1.93%                        |

**Table 1.** Description of the dataset.

measurements, positive EmA test, and celiac-type human leukocyte antigen haplotypes<sup>6,16</sup>. Recently, such diagnostic approach has been shown to be applicable also in adults<sup>17</sup>. It highlights the future importance of EmA testing and calls for solutions related to its labor-intensiveness and subjectivity.

The aim of this research was to determine whether supervised machine learning can be applied to create an automated method with expert comparable precision for the assessment and classification of the IgA-class EmA test for celiac disease.

## Results

In 2017-2018, in celiac disease service laboratory at the Tampere University, Tampere, Finland, 2597 high-quality images of IgA-class EmA test samples were obtained. All of them were classified as positive, negative, IgA deficient, or equivocal (Fig. 1) by an expert evaluator as presented in Table 1 and included in the study.

Machine learning was deployed to create a new method of automatic assessment and classification of the IgA-class EmA test for celiac disease. The classification was based in AdaBoost with SVM, and the sample features obtained through multi-scale, rotational invariant, co-occurrence among adjacent local binary patterns. Two

|                   |                    |                 |              |                    |               |
|-------------------|--------------------|-----------------|--------------|--------------------|---------------|
| Expert evaluation | I. Positive        | 195             | 77           |                    | 2             |
|                   | II. Negative       | 17              | 2236         | 3                  | 4             |
|                   | III. IgA deficient | 1               | 6            | 6                  |               |
|                   | IV. Equivocal      | 12              | 32           |                    | 6             |
|                   |                    | I. Positive     | II. Negative | III. IgA deficient | IV. Equivocal |
|                   |                    | Predicted class |              |                    |               |

**Figure 2.** Confusion matrix of the classification model.

| Measure of performance    | Model 1 | Model 2 |
|---------------------------|---------|---------|
| Accuracy                  | 96.80%  | 98.85%  |
| Classification error      | 3.20%   | 1.15%   |
| Sensitivity               | 82.84%  | 98.91%  |
| Specificity               | 99.40%  | 98.81%  |
| F <sub>1</sub> score      | 0.65    | 0.75    |
| Cohen's kappa coefficient | 0.85    | 0.98    |

**Table 2.** Analysis of performance of Model 1 (based on the whole sample size) and Model 2 (supplemental, randomly under-sampled).

SVM models were trained - Model 1 based on the whole sample-size ( $n = 2597$ ) and supplemental Model 2 based on the randomly under-sampled size.

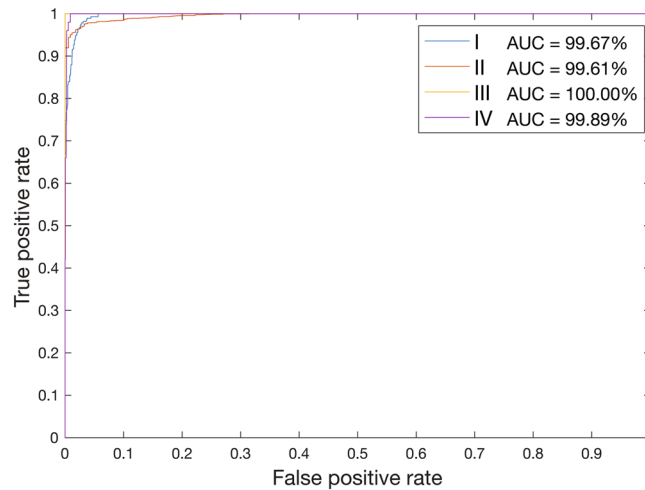
From 274 samples graded as positive by the expert evaluator, Model 1 predicted 195 to be positive, 77 negative, and 2 equivocal (Fig. 2). From 2260 samples graded as negative by the expert evaluator, the algorithm predicted 17 to be positive, 2236 negative, 3 as IgA deficient, and 4 as equivocal. Out of 13 samples classified by the expert as IgA deficient, 6 were classified by the algorithm as IgA deficient, 1 as positive, and 6 as negative. Finally, from the 50 equivocal samples defined by the expert evaluator, 12 were predicted to be positive, 32 negative, and 6 as equivocal.

The Model 1 developed in this research was characterized by accuracy of 96.80% and classification error of 3.20% (Table 2). The sensitivity was 82.84% and specificity 99.40%. The F<sub>1</sub> score was 0.65 and Cohen's kappa coefficient was 0.85. The AUC was 99.67%, 99.61%, 100%, and 99.89%, for I, II, III, and IV class, respectively (Fig. 3). The mean assessment time per image was 16.11 seconds (standard deviation of 0.52 seconds).

To control for the possibility of multi-class classification problems supplemental Model 2 was created, deploying random under-sampling. The measures of performance of Model 2 were remarkably similar to the ones obtained with Model 1 (Table 2). The Model's 2 accuracy was 98.85%, and the error rate was 1.15%. Its sensitivity and specificity were 98.91% and 98.81%, respectively. The F<sub>1</sub> score was 0.75 and Cohen's kappa coefficient was 0.98.

## Discussion

**Deployed machine learning methods.** In this study, we selected SVMs for the development of automatic classification method for IgA-class EmA test, as they represent a class of non-probabilistic binary linear classifiers, able to map their inputs into high-dimensional feature spaces. They are a popular machine learning technique that has been applied successfully in other fields of biomedicine<sup>18–24</sup>. In this study, a feature descriptor, called multiscale co-occurrence among adjacent local binary patterns after edge enhancement, was applied. The descriptor is multi-scale, which means it is able to cope with different sizes of samples' structures such as different vessels' diameter, reticulin fibers' length, and Wharton's jelly conformation. Moreover, the descriptor is rotation invariant, allowing for a precise classification independently of the slides' orientation in the microscope. Besides, it has a high descriptive power, which enables detecting subtle differences between samples and consequently their correct classification. The power of the descriptor results from the auto-correlation matrix generated by the multiple local binary patterns mapping. Such mapping enables adapting to non-uniform illumination conditions. Finally, due to the expansion of the canonical local binary pattern premise of center-based texture description to the inter-pattern correlation, the descriptor is dependent only on the contrast of magnitude between the central pattern and its surrounding patterns<sup>25</sup>.



**Figure 3.** Receiver operating characteristic curves for the classification model.

The high variability between the number of samples in the four different groups might seem the main limitation of the study. However, when applying machine learning to clinical sciences, where usually classes are not represented equally, imbalanced data is a typical problem. Nevertheless, it does not mean that machine learning cannot be applied to clinical problems. In the case of the presented Model 1, one might expect to encounter multi-class classification problems. Datasets like ours, that represent a cross-section of subjects suspected for celiac disease, are typically imbalanced, which reflects the sample distribution in a genuine clinical setting. The vast majority of the patients will be in the “Negative” class and a very small minority will be in the “Positive” class.

To address the possibility of multi-class classification problems we used several techniques. First of all, AdaBoost ensemble technique was used to combine weak learners to create a strong learner that can make accurate predictions. AdaBoost deals with the class imbalance problem by maintaining a set of weights on the training dataset in the learning process. Moreover, 10-fold cross-validation provides positive class samples with a higher chance of inclusion in the training and testing phases. Furthermore, while presenting the results not only accuracy was showed, but also different performance metrics that are regarded as more correct in case of imbalanced datasets, i.e. a confusion matrix and ROC curves. Finally, a supplementary Model 2 was created, that deployed a sample obtained after random under-sampling without replacement.

In the case of the “IgA-deficient” class, representing only 0.50% of the dataset, overfitting was observed (AUC = 100%). However, as the model’s performance is reported by the AUC, which is computed according to a one-versus-all approach for multi-class problems, the obtained results are independent between classes. Possible future expansion of the presented research would be by increasing the samples’ number and balancing the samples distribution between classes.

**Limitations and strengths of the study.** The developed method is dependent on the quality of the images that compose the dataset, which is one of the most common issues in the image-based machine learning field. Therefore, imaging parameters’ standardization and acquisition area that characterized the used dataset were crucial for the correct classification of the samples.

The principal advantage of the study is the large sample size ( $n = 2597$  samples) that was used to create the classification model which represents the actual distribution of the EmA samples in clinical practice. Furthermore, the model was trained with the results of the evaluation made beforehand by a highly qualified specialist with over 20 years of experience in the assessment of IgA-class EmA test for celiac disease. Although only one expert graded the EmA images, this follows a quality-controlled standard procedure fulfilling European guidelines. Thus, the results show that machine learning can be used to provide similar precision as expert evaluation.

**Significance of the findings and suggestions for future research.** The presented findings’ combination provides support for the conceptual premise that deploying machine learning for automatic classification of IgA-class EmA test in celiac disease is possible. These results suggest that creating a model characterized by both satisfying sensitivity and specificity is conceivable. To the knowledge of the authors, the method is the first fully-automated, user-independent algorithm deploying machine learning for IgA-class EmA test for celiac disease assessment published in the literature.

This study lays the groundwork for future research into the application of artificial intelligence modalities in the field of celiac disease diagnostics. Its empirical findings demonstrate that implementation of machine learning enables quick and precise EmA test analysis comparable to an expert-conducted evaluation and provides an automated method in the diagnostic work-up in addition to common TG2-Ab ELISA test. This approach could be extended by deploying bigger sample size and multiple expert sample evaluators.

The presented findings have important implications for expanding the use of IgA-EmA tests by healthcare professionals for celiac disease diagnostics. The use of machine learning to provide automatic analysis of IgA-EmA tests can facilitate work both in centers with a low incidence of celiac disease and in low-resource settings with



little access to specialists able to provide a high-quality evaluation. Furthermore, deployment of machine learning algorithms would make the IgA-EmA test assessment less time-consuming and more cost-effective.

## Conclusions

In conclusion, in this study, a new method of the automatic analysis of IgA-class EmA test for celiac disease deploying machine learning was presented. It is the first report on using machine learning for automatic classification of IgA-class EmA tests for celiac disease. The model is characterized by the sensitivity of 82.84% and specificity of 99.40%. The mean assessment time is 16.11 seconds per sample. The study confirmed the possibility of machine learning application for automatic evaluation and classification of the IgA-class EmA test for celiac disease.

## Materials and Methods

**Study material.** The study material included consecutive serum samples collected due to clinical suspicion of celiac disease and analyzed for EmA during the 2017-2018 period in celiac disease service laboratory at the Tampere University, Tampere, Finland. The laboratory serves as a reference entity and takes part in regular international external quality-control assessment by the United Kingdom National External Quality Assessment Scheme. The EmA test was standardized according to European standardization working group (ESPGHAN and European Medical Research Councils Clinical Network for Gastroenterological Immunology: Serological Screening for Celiac Disease)<sup>26</sup>.

The EmA testing was part of clinical studies approved by the Regional Ethics Committee of Tampere University Hospital District. In the above-mentioned studies, all study participants provided written informed consent.

**Determination of IgA-class EmA.** The IgA-class EmA testing was carried out according to the protocol developed by the European Working Group on Serological Screening for Celiac Disease<sup>26</sup>. Five  $\mu\text{m}$  thick, full transverse human umbilical cord cryostat sections were placed on Vectabond<sup>TM</sup> (Vector Laboratories, Burlingame, CA, USA)-pretreated slides. Serum samples were initially diluted 1:5 in phosphate-buffered saline (PBS, pH 7.4). The slides were incubated for 30 minutes with diluted serum samples at room temperature. Afterwards, they were washed in PBS (pH 7.4) and incubated with fluorescein isothiocyanate-conjugated rabbit F(ab')<sub>2</sub> anti-human IgA (dilution 1:160, DAKO A/S, Glostrup, Denmark) for 30 minutes at room temperature. After subsequent washing and mounting, the slides were analyzed with a fluorescence microscope (Olympus BX60, UPLANFL 20x objective). Positivity was defined as fluorescence of reticulin fibers in the vascular wall, external border of the vascular wall, and structures of Wharton's jelly. All positive and equivocal sera were further diluted up to 1:4000 (1:5, 1:50, 1:100, 1:200, 1:500, 1:1000, 1:4000).

Following a quality-controlled standard procedure, all of the slides were classified by a specialist, with more than 20 years of experience in evaluation of IgA-class EmA, into one of the following four classes: I - positive, II - negative, III - IgA deficient, and IV - equivocal (Fig. 1).

The samples were photographed with pre-set camera settings (gamma 1.08, saturation 0.20, blue and red channels were excluded, green channel 7.04, acquisition time 1.00 s). The images were exported in .jpg format. To create the algorithm, a dataset of 2597 high-quality IgA-class EmA images was used. The above number of samples is the result of the collection of data on the diagnosis of patients with suspected celiac disease in a university hospital during the period of 22 months.

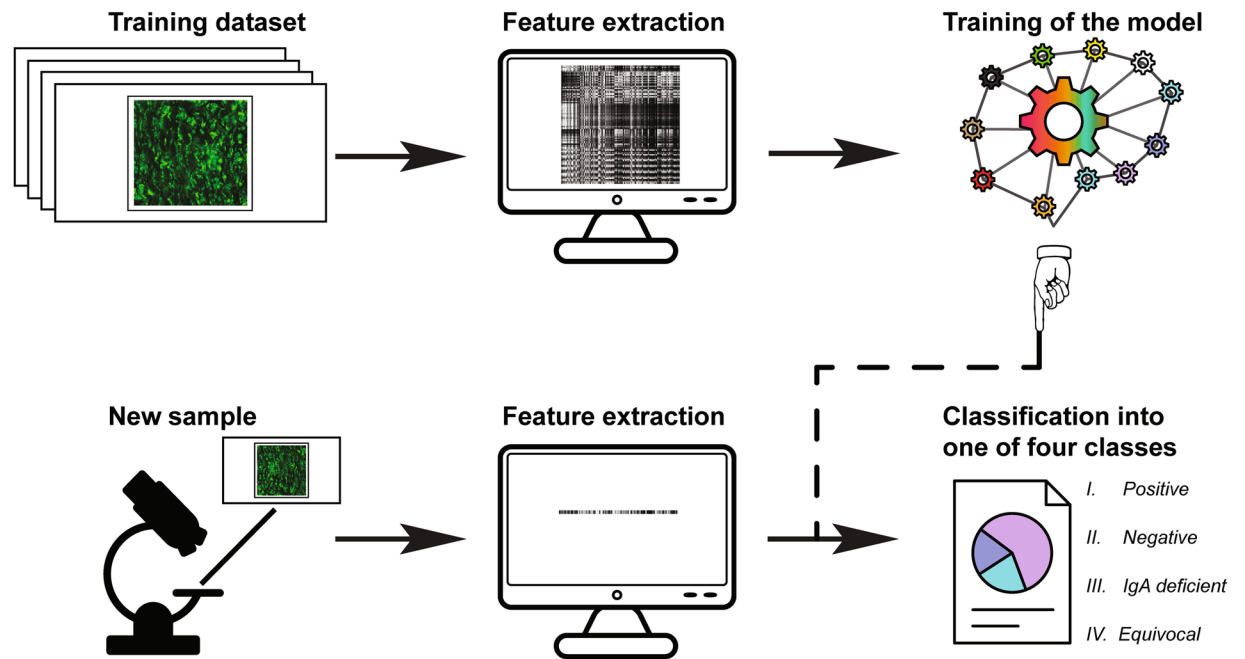
**Development of a classification model.** Conventionally, development of a classification model is a three-stage process (Fig. 4). During the first stage features of the images are extracted using feature descriptors. Next, the model is trained and tested. Finally, the performance of the created model is evaluated.

A feature descriptor proposed by Nosaka *et al.*<sup>25</sup> was applied consecutively to all of the *a priori* blinded images from the dataset, to extract the feature set. After enhancing edges, the descriptor extracts the co-occurrence among adjacent local binary patterns with three search radiuses 1, 2, and 4 pixels. These provide multi-scale rotation invariance and high descriptive ability.

The previously prepared feature set was used to train error-correcting output codes multi-class support vector machine model (SVM) (with a radial basis function kernel, as a one-versus-all classifier, ten-fold cross-validated). In other words, given labeled training data, the SVM model, where a data point is viewed as a p-dimensional vector, was used to separate such points with (p-1) dimensional hyperplane and provide an optimal hyperplane which categorizes new examples.

In each training-testing fold, the dataset was divided randomly as following: 70% for training and 30% for testing. Every time, a one-versus-all SVM was trained for each class to prevent overfitting even in cases where the length of the feature vector was larger than the number of observations. To prevent overfitting, SVMs use regularization. It is done by applying non-linear kernels and tuning of the kernels and regularization parameters. Both tuning and regularization parameters are optimized by multiple (in this study 10-fold) and consecutive cross-validated training-and-testing. This decreases the generalization error, dependent on the margin (distance between class centers) but independent of the feature space<sup>27</sup>.

After mapping the feature vector into a complex hyperspace, to minimize exponential loss, the parameters were tuned using AdaBoost, a machine learning meta-algorithm resistant to overfitting<sup>28</sup>, with one hundred consecutive learning cycles. AdaBoost is based on a cycle of consecutive training and tuning of weak trainers' set. After each iteration the weak trainers and their weights are tweaked to optimize the separation between classes. The weak trainers that misclassified a sample are discarded and replaced by new ones, with random parameters. Through such an evolution of many generations of weak learners, AdaBoost provides a classification method less prone to overfitting<sup>29</sup>. To control for ascertainment bias, decision trees (weak trainer models) with ten surrogate splits at each branch node were applied in the present methodology. With each AdaBoost iteration, the branches of the decision trees were pruned, and the weights recalculated, improving the classification performance.



**Figure 4.** Development and usage of the classification model.

In this study, two SVM models were trained. Model 1 was based on the whole sample size ( $n = 2597$ ). Supplemental Model 2 was created to adjust the class distribution of a dataset and address the possibility of multi-class classification problems. Considering the dataset's characteristics, Model 2 was created deploying random under-sampling. Samples from the majority class (negative) were randomly removed without replacement until the number of samples in negative class and the positive class (according to expert evaluation) was even.

Both of the classification models were developed and tested using MATLAB<sup>®</sup> (version R2018b, *Image Processing Toolbox, Statistics and Machine Learning Toolbox*).

**Model evaluation.** To evaluate the performance of the models a confusion matrix was created, and the following parameters were calculated: classification error (using 10-fold cross-validation), sensitivity, specificity, accuracy, and the receiver operating characteristic curves for the calculation of the area under the curve (AUC). The performance-related information of a classifier was obtained creating a *classperf* object by using the *classperf* function (MATLAB<sup>®</sup>; version R2018b). The accuracy, which defines the closeness of a measured value to an expert-evaluation value, is a positive scalar defined as the number of correctly classified samples divided by the number of classified samples, where inconclusive results are not counted<sup>30</sup>. The error rate of the classifier is a positive scalar, defined as the number of incorrectly classified samples divided by the number of classified samples<sup>30</sup>. The sensitivity of the classifier is a positive scalar, defined as the number of correctly classified positive samples divided by the number of true positive samples<sup>30</sup>. The specificity of the classifier is a positive scalar, defined as the number of correctly classified negative samples divided by the number of true negative samples<sup>30</sup>.  $F_1$  score is a positive scalar, defined as the harmonic average of precision (samples correctly classified as positive out of all positive classified samples) and sensitivity<sup>31</sup>. Cohen's kappa coefficient is a positive scalar defined as the quotient of the difference between the agreement rate and the hypothetical probability of chance agreement, and the complement of the hypothetical probability of chance agreement<sup>32</sup>. The AUCs were obtained deploying the *perfcurve* function (MATLAB<sup>®</sup>; version R2018b). Finally, mean analysis time per image was calculated.

### Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request and subject to the ethical approvals in place and material transfer agreements.

### References

1. Lindfors, K. *et al.* Coeliac disease. *Nat Rev Dis Primers*. **5**(1), 3 (2019).
2. Singh, P. *et al.* Global prevalence of celiac disease: Systematic Review and Meta-analysis. *Clin Gastroenterol Hepatol*. **16**(6), 823–36.e2 (2018).
3. Murray, J. A., Frey, M. R. & Oliva-Hemker, M. Celiac disease. *Gastroenterology*. **154**(8), 2005–8 (2018).
4. Kelly, C. P., Bai, J. C., Liu, E. & Leffler, D. A. Advances in diagnosis and management of celiac disease. *Gastroenterology*. **148**(6), 1175–86 (2015).
5. Taavela, J. *et al.* Validation of morphometric analyses of small-intestinal biopsy readouts in celiac disease. *PLoS One*. **8**(10), e76163 (2013).
6. Husby, S. *et al.* European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr*. **54**(1), 136–60 (2012).

7. Marsh, M. N. Grains of truth: evolutionary changes in small intestinal mucosa in response to environmental antigen challenge. *Gut*. **31**(1), 111–4 (1990).
8. Biagi, F. *et al.* The prevalence and the causes of minimal intestinal lesions in patients complaining of symptoms suggestive of enteropathy: a follow-up study. *J Clin Pathol*. **61**(10), 1116–8 (2008).
9. Kakar, S., Nehra, V., Murray, J. A., Dayharsh, G. A. & Burgart, L. J. Significance of intraepithelial lymphocytosis in small bowel biopsy samples with normal mucosal architecture. *Am J Gastroenterol*. **98**(9), 2027–33 (2003).
10. Corazza, G. R. & Villanacci, V. Coeliac disease. *J Clinical Pathol*. **58**(6), 573–4 (2005).
11. Ravelli, A., Bolognini, S., Gambarotti, M. & Villanacci, V. Variability of histologic lesions in relation to biopsy site in gluten-sensitive enteropathy. *Am J Gastroenterol*. **100**(1), 177–85 (2005).
12. Bonamico, M. *et al.* Duodenal bulb biopsies in celiac disease: a multicenter study. *J Pediatr Gastroenterol Nutr*. **47**(5), 618–22 (2008).
13. Rashid, M. & MacDonald, A. Importance of duodenal bulb biopsies in children for diagnosis of celiac disease in clinical practice. *BMC Gastroenterol*. **9**, 78 (2009).
14. Weir, D. C., Glickman, J. N., Roiff, T., Valim, C. & Leichtner, A. M. Variability of histopathological changes in childhood celiac disease. *Am J Gastroenterol*. **105**(1), 207–12 (2010).
15. Corazza, G. R. *et al.* Comparison of the interobserver reproducibility with different histologic criteria used in celiac disease. *Clin Gastroenterol Hepatol*. **5**(7), 838–43 (2007).
16. Werkstetter, K. J. *et al.* Accuracy in diagnosis of celiac disease without biopsies in clinical practice. *Gastroenterology*. **153**(4), 924–35 (2017).
17. Fuchs, V. *et al.* Serology-based criteria for adult coeliac disease have excellent accuracy across the range of pre-test probabilities. *Aliment Pharmacol Ther*. **49**(3), 277–284 (2018).
18. Zhu, X.-J., Feng, C.-Q., Lai, H.-Y., Chen, W. & Hao, L. Predicting protein structural classes for low-similarity sequences by evaluating different features. *Knowl-Based Syst*. **163**, 787–93 (2019).
19. Chen, W., Lv, H., Nie, F. & Lin, H. i6mA-Pred: Identifying DNA N6-methyladenine sites in the rice genome. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btz015>. [Epub ahead of print] (2019).
20. Zuo, Y. *et al.* iDPF-PseRAAAC: A Web-Server for Identifying the Defensin Peptide Family and Subfamily Using Pseudo Reduced Amino Acid Alphabet Composition. *PLOS ONE*. **10**(12), e0145541 (2015).
21. Tang, H. *et al.* HBPred: a tool to identify growth hormone-binding proteins. *Int J Biol Sci*. **14**(8), 957–64 (2018).
22. Zuo, Y.-C. *et al.* Discrimination of membrane transporter protein types using K-nearest neighbor method derived from the similarity distance of total diversity measure. *Molecular Omics*. **11**(3), 950–7 (2015).
23. Feng, C. Q. *et al.* iTerm-PseKNC: a sequence-based tool for predicting bacterial transcriptional terminators. *Bioinformatics*. **35**(9), 1469–77 (2019).
24. Zuo, Y. C. *et al.* Predicting peroxidase subcellular location by hybridizing different descriptors of Chou' pseudo amino acid patterns. *Anal Biochem*. **458**, 14–9 (2014).
25. Nosaka, R., Suryanto, C. H. & Fukui, K. editors. Rotation invariant co-occurrence among adjacent LBPs, Berlin, Heidelberg: Springer Berlin Heidelberg (2013).
26. Stern, M. Comparative evaluation of serologic tests for celiac disease: a European initiative toward standardization. *J Pediatr Gastroenterol Nutr*. **31**(5), 513–9 (2000).
27. Han, H. & Jiang, X. Overcome support vector machine diagnosis overfitting. *Cancer Inform*. **13**(Suppl 1), 145–58 (2014).
28. Freund, Y. & Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*. **55**(1), 119–39 (1997).
29. Schapire, R. E. Explaining AdaBoost. In: Schölkopf, B., Luo, Z. & Vovk, V., editors. Empirical inference: Festschrift in honor of Vladimir N Vapnik. Berlin, Heidelberg: Springer Berlin Heidelberg; p. 37–52 (2013).
30. Altman, D. Practical Statistics for Medical Research. Chapman&Hall/CRC (2006).
31. Chinchor, N. MUC-4 Evaluation Metrics. Proc. of the 4<sup>th</sup> Conference on Message Understanding Conference; 22–29 (1992).
32. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ Psych Meas*. **20**(1), 37–46 (1960).

## Acknowledgements

This study was supported by the Academy of Finland, the Sigrid Juselius Foundation, and the Competitive State Research Financing of the Expert Area of Tampere University Hospital.

## Author Contributions

Study concept and design: Florentino Luciano Caetano dos Santos, Katri Kaukinen, Jari Hyttinen, Katri Lindfors. Acquisition of data: Kaija Laurila, Katri Kaukinen. Analysis and interpretation of data: Florentino Luciano Caetano dos Santos, Irmina Maria Michalek, Kaija Laurila. Drafting the manuscript: Florentino Luciano Caetano dos Santos, Irmina Maria Michalek, Katri Lindfors. Critical revision of the manuscript for important intellectual content: Florentino Luciano Caetano dos Santos, Irmina Maria Michalek, Kaija Laurila, Katri Kaukinen, Jari Hyttinen, Katri Lindfors. Statistical analysis: Florentino Luciano Caetano dos Santos. Obtained funding: Katri Kaukinen, Jari Hyttinen, Katri Lindfors. Technical and material support: Florentino Luciano Caetano dos Santos, Irmina Maria Michalek. Study supervision: Katri Kaukinen, Jari Hyttinen, Katri Lindfors.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019