Enhancing patient stratification and interpretability through class-contrastive and feature attribution techniques

1st Sharday Olowu
Department of Computer Science
University of Cambridge
Cambridge, UK
sylo2@cantab.ac.uk

2nd Neil Lawrence
Department of Computer Science
University of Cambridge
Cambridge, UK
ndl21@cam.ac.uk

3rd Soumya Banerjee
Department of Computer Science
University of Cambridge
Cambridge, UK
https://orcid.org/0000-0001-7748-9885

***Abstract*—A crucial component of treating genetic disorders is identifying the genes and gene modules that drive disease processes. While Next-Generation Sequencing (NGS) provides rich data for this task, current machine learning approaches often lack explainability and fail to account for gene correlations. We develop a comprehensive framework of machine learning techniques for explainable patient stratification in inflammatory bowel disease, focusing on Crohn's disease (CD) subtypes: CD with deep ulcer, CD without deep ulcer and IBD-controls. Our approach combines Gaussian Mixture Modelling for patient stratification, a modified kernelSHAP algorithm accounting for gene co-expression, systematic identification of gene modules, and class-contrastive analysis for explaining individual patient phenotypes. This framework confirms known disease-associated genes while unveiling novel genetic factors potentially underlying CD heterogeneity. Gene Ontology enrichment analysis validates the biological relevance of identified gene modules and associated pathways. Our methods offer a versatile toolkit for analysing high-dimensional, correlated biological data across diverse disease contexts.**

***Index Terms*—Explainable AI.**

I. Introduction

The wealth of RNA-Seq data from Next-Generation Sequencing has created new opportunities for analysing the genetic basis of disease, but current machine learning approaches often lack interpretability and overlook gene co-expression patterns. We introduce an explainable machine learning framework for uncovering the genetic aetiology of Crohn's disease (CD) subtypes, accounting for key gene correlation patterns. Our approach is summarised in Figure 1. All code is available from the following repository: https://zenodo.org/doi/10.5281/zenodo.10278383 and Supplementary Material for this work can be accessed at https://osf.io/efz95.

II. Related work

A. Explainability

1) Class-contrastive techniques: Class-contrastive techniques can be used to find the impact of particular features on the model output. This can be very useful for improving the transparency of a model. For example, Banerjee et al. generate explainability for mortality predictions of patients in [3], as predicted by deep learning models. By setting the presence or absence of binary features such as "suffering with depression" or "lack of family support", they were able to find the isolated effect of each feature on the risk of patient mortality (as predicted by a black-box model). This technique is usually used with categorical features. The class-contrastive approach has been used for self-supervised clustering of RNA-Seq data [4], but not for generating explanations of disease subtype based on genomic data. In this work, we extend the class-contrastive approach to demonstrate the impact of gene modules on disease subtype, taking into account the underlying gene expression distributions.

2) Applications of SHAP: SHapley Additive exPlanations (SHAP) [5] is another state-of-the-art explainability technique. Fast approximations of SHAP have been applied to analyse gene expression data [6]–[10], such as kernelExplainer, treeExplainer and gradientExplainer [11]. Yu et al. use a deep autoencoder [8] to learn gene expression representations, applying treeExplainer SHAP to measure the contributions of genes to each of the latent variables.

Although SHAP has shown success in this line of work, one significant problem is that it assumes feature independence. This means that when applying SHAP to find the contribution of genes, it is assumed that there are no correlations between genes. This is unrealistic because genes are often correlated and/or regulated by other genes; this is governed by complex gene regulatory networks [12].

There have been some attempts to include feature dependence, for example in the linearExplainer and treeExplainer [13] SHAP variants. However, linear models are not suitable for modelling complex gene-gene and genotype-phenotype relationships. The treeExplainer is also limited to tree-based ensemble methods, which can be difficult to visualise and interpret. We aim to address this by incorporating inter-feature dependence for kernelSHAP (kernelExplainer), which is model-agnostic and therefore applicable in many more contexts. By incorporating inter-feature dependence, we can more accurately identify potential genes, gene modules, and associated biological pathways implicated in disease processes.

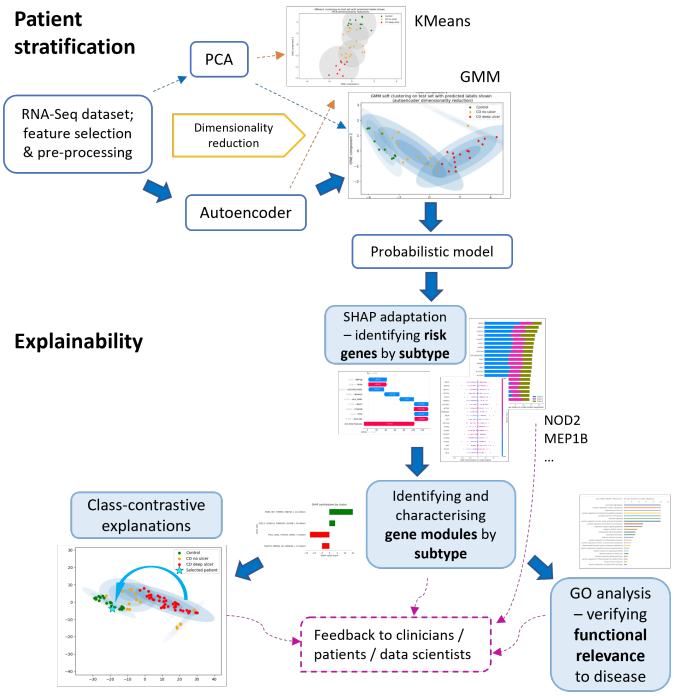

Fig. 1: Overview of our computational framework. We identify genes and gene modules implicated in Crohn's disease (CD) subtypes. Starting with RNA-Seq data, reduce dimensionality using PCA and an autoencoder (adapted from [1]). We then use a probabilistic model (GMM) and post-processing to cluster and classify patients into disease subtypes (CD with deep ulcer, CD without deep ulcer and control). In order to explain our model we develop: 1. an extension of Shapley Additive Explanations (SHAP) to account for gene correlations, and 2. a class-contrastive technique to visually demonstrate the effect of changing gene expression on individual patients. Potential gene modules are identified using novel data integration and WECR clustering [2]. We confirm our findings by referencing peer-reviewed studies and conducting a Gene Ontology (GO) enrichment analysis.

Aas et al. propose to achieve inter-feature dependence for kernelSHAP in [14]. Here, a multivariate Gaussian distribution is constructed using a sample mean vector and covariance matrix, calculated from the training data. For each input instance (corresponding to a coalition of features), this model is updated under a Bayesian framework and used to generate synthetic samples for the calculation of conditional expectations required by the kernelSHAP algorithm. In the context of RNA-Seq analysis, this method is not appropriate because the model of relationships between genes can differ significantly between input instances. More specifically, the Bayesian framework leads to the modification of feature correlations and the means of marginalised features to varying degrees.

In this work, we use an alternative approach to address the need to take account of consistent relationships between features. We construct a multivariate Gaussian distribution to model these relationships. However, when updating this model between input instances, we only modify the mean and variance of those features present in the current model's coalition, leaving feature correlations intact. In this way, we preserve our knowledge of the underlying dataset, including the relationships between genes originally captured in the training data. This results in a more representative multivariate Gaussian distribution. Since we are aiming to draw insights about consistent relationships between genes, this promotes more realistic SHAP values and therefore cluster explanations.

B. *Cluster analysis and gene module identification*

Identifying gene modules is a crucial step in characterising the genetic component of disease. Current techniques tend to include a clustering aspect and/or network construction [15]–[17] to organise genes, such as Weighted Gene Co-expression Network Analysis [18], [19]. However, they can be sensitive to noise, with high computational complexity that can limit scalability to larger datasets. Our approach also uses clustering, but reduces complexity and the impact of noise by implicitly capturing gene and sample relationships. We achieve this by using Gaussian Mixture Modelling and a deep autoencoder that can infer both linear and non-linear relationships. We adapt our mixture-based clustering model for classifying disease subtype based on RNA-Seq data.

Our approach explicitly accounts for inter-feature dependence by analysing the underlying data distributions and correlations between genes, using data from real patients.

III. Data and Methods

A. *Patient stratification*

The study used a publicly available transcriptomic dataset [20], [21] containing RNA-Seq data from ileal tissue samples of paediatric subjects. The subjects were divided into subtypes: CD with deep ulcer, CD without deep ulcer and IBD-controls. Dimensionality reduction was performed using an autoencoder and PCA, followed by clustering with Gaussian Mixture Models (GMMs) and K-means. While these unsupervised techniques were used to identify potential clusters of patients with similar transcriptomic characteristics, the process was integrated into a semi-supervised framework by applying a novel post-processing algorithm that maps clusters to patient phenotypes. K-means provided straightforward cluster assignments, whereas GMMs enabled a probabilistic model, allowing for a more sophisticated classification of patients into CD subtypes. Please see Supp. Material Section 4.1 for more essential details.

B. *Clustering explainability*

1) Modifying kernelSHAP to identify risk genes: Feature importance ranking provides crucial insights into machine learning model predictions. We developed an extension of kernelSHAP, a leading algorithm for this task, to identify genes most influential in predicting CD subtypes of patients from RNA-Seq data, which we demonstrate using Gaussian Mixture Models. While the original kernelSHAP [5] assumes feature

independence, our method accounts for the correlated nature of gene expression by incorporating inter-feature dependence. Building upon work by Aas et al. [14], we approximate conditional feature distributions using a multivariate Gaussian distribution derived from training data, adjusting for each input instance to maintain gene relationships across coalition scenarios (see Supp. Material Section 4.2.1). This generates more realistic synthetic samples, yielding more accurate SHAP values and reliable explanations of disease subtypes at both patient and cluster levels. In practice, this can enable the systematic prioritisation of genes for further analysis, based on their predicted influence on CD subtype presentation.

2) Identification and characterisation of potential gene modules: We introduce a novel approach to identify disease-related gene modules by integrating the resulting SHAP values (from the previous section) with the gene expression (RNA-Seq) data. The main idea is to use both gene activity level *and* SHAP gene importance to determine those genes which are likely to be working in concert as "modules" wrt. disease processes. The integration formula is as follows: $v_{pg} = x_{pg}\frac{\sum_i abs(s_{ig})c_i}{n}$ where v_{pg} represents the integrated value for patient p and gene g, x_{pg} is the gene expression, s_{ig} is the SHAP value, c_i indicates membership in a disease subtype, and n is the total number of patients in that subtype. This formula uniquely combines gene importance (via SHAP values) with expression levels, providing a more comprehensive view of gene influence on disease subtypes. The integrated data undergoes Weighted Ensemble Consensus of Random (WECR) K-Means clustering [2], with the optimal number of clusters determined using four validation metrics: Bayesian Information Criterion, Davies-Bouldin Index, Silhouette Score, and Calinski-Harabasz Index. We also demonstrate a technique for characterising the resulting gene modules by a sum of SHAP values across each gene module for a given disease subtype, as shown in Figure 3, to indicate the type and magnitude of influence of each identified module. The results are verified through Gene Ontology enrichment analysis, offering insights into the functional relevance of each gene module wrt. specific biological pathways, and potential impacts on disease subtype. Further essential details of the approach can be found in Supp. Material Section 4.2.2.

3) Class-contrastive technique for patient-specific explainability: A novel class-contrastive technique was developed to explain clusters specific to a patient by generating explanations that provide a contrast to another class.

For patients with Crohn's Disease (CD), class-contrastive explanations can be generated by artificially modifying the expression of genes in a given module to more closely resemble those of control subjects. We achieved this by assigning a new expression value (v) for each chosen gene, calculated as the mean value for the expression of this gene across all control individuals, as shown in the equation: $\forall g \in G, \quad v_{pg} = \frac{1}{N}\sum_i c_i x_{ig}$, where x represents an expression value, p is the selected patient, g is the selected gene, and G is the full set of genes in the module. The summation is performed over all control individuals i, where c_i serves as the indicator variable for the control group and N represents the total number of control individuals. After modifying these expression values, the patient's data is rerun through the clustering model to observe whether their cluster assignment changes - for instance, whether a patient originally clustered in the 'severe disease' group might shift to a 'mild disease' or even 'healthy control' cluster, thereby revealing the importance of those modified genes in disease classification. More comprehensive details can be found in Supp. Material Section 4.2.3.

IV. Results and Discussion

A. Gaussian Mixture Model (GMM) and KMeans clustering

GMM clustering with autoencoder and tSNE outperformed other methods in stratifying and classifying patients into Chrohn's disease subtypes, as shown in Table I. This approach was selected for downstream analysis. More details are in Supp. Material Section 5.1.

TABLE I: Clustering and classification evaluation results for novel classifiers based on Gaussian Mixture Model (GMM) and KMeans models, using autoencoder and PCA dimensionality reduction methods. Results shown for binary classification (controls and all CD patients) and multi-class classification (control, CD no ulcer and CD deep ulcer) of disease subtype.

		Binary (control & CD)		**Multi-class (all labels)**	
		Autoencoder	PCA	Autoencoder	PCA
GMM	Accuracy / %	94.9	92.3	71.8	64.1
	F1-Score / %	96.7	94.9	71.5	62.6
	Silh. score	0.382	0.410	0.320	0.317
KMeans	Accuracy / %	84.6	82.1	64.1	59.0
	F1-Score / %	89.3	88.1	61.9	58.3
	Silh. score	0.556	0.409	0.469	0.334

B. Cluster explanation using kernelSHAP adapted for feature dependence

We coupled our GMMs to kernelSHAP [5] to generate explainability for each cluster, including visualisations [22]. Our modification of the kernelSHAP method incorporates feature dependence to more accurately model gene correlations and regulatory relationships, enabling identification of the most influential genes in predicting disease subtypes. Figure 2 shows a summary of the top 20 most influential genes, with bars depicting their influence on "CD deep ulcer" (blue), "CD no ulcer" (pink) and "control" (green) classifications.

Compared to the original kernelSHAP method (Supp. Material Fig. 6), where 4 of the top 5 genes (BPIFB1, FOXD1, C19orf59 and NAT8B) had no known IBD associations, our SHAP adaptation ranked established IBD genes significantly higher, with NOD2, MEP1B, and FOLH1 appearing in the top 5. These are examples of known susceptibility genes for IBD. For example, NOD2 plays a crucial role in bacterial sensing by recognising muramyl dipeptide on bacterial cell envelopes, leading to oligomerisation and RICK kinase

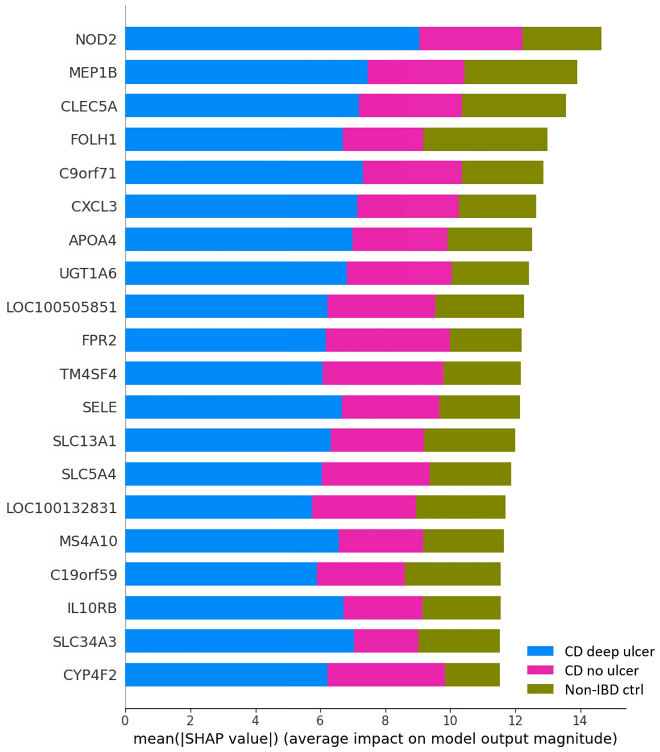

Fig. 2: Summary plot showing top 20 genes in terms of their average impact on class predictions across all patients. This is for a model which accounts for feature dependence. The blue, pink and green bars depict the magnitude of influence of a gene on the "CD deep ulcer", "CD no ulcer" and "control" classes respectively. The greatest proportion of influence of genes is attributed to the "CD deep ulcer" cluster. Genes like NOD2, MEP1B and FOLH1 are within the top 5 and known as susceptibility genes for IBD. The most significant gene identified overall is NOD2, which is known to be strongly associated with IBD.

binding. This activates the NF-κB pathway [23], [24], resulting in pro-inflammatory cytokine accumulation and tissue inflammation [25]. Our method also identified additional IBD-implicated genes not detected by the original algorithm, including IL10RB, CXCL3, APOA4, SLC13A1 and SLC5A4. Chemokines such as CXCL3 and cytokines such as IL10 are associated with inflammatory processes in IBD [26]. For example, Interleukin-10 maintains intestinal haemostasis by suppressing pro-inflammatory cytokines like TNF and IL-12, with its disruption leading to IBD symptoms [27]. Interestingly, our analysis also highlighted two uncharacterised "LOC" genes [28] that may represent novel therapeutic targets, although further validation is needed, as some identified genes such as SELE currently show limited evidence of IBD association.

Compared to state-of-the-art approaches that commonly apply SHAP to neural networks [6]–[9], our GMM-based probabilistic model coupled with the kernelSHAP adaptation provides improved interpretability while accounting for gene dependencies. The method generates both local explanations for individual patients and global insights across the dataset, with the incorporation of inter-feature dependence resulting in improved alignment with established IBD literature. Please see Supp. Material Section 5.2 for more details.

C. Identification and characterisation of potential gene modules

To identify disease-related gene modules for each CD subtype, our novel approach integrates SHAP values with gene expression data, before applying Weighted Ensemble Consensus of Random K-Means clustering [2]. In analysing the CD deep ulcer subtype, for example, it reveals four modules containing known IBD genes, like IRGM [29], CXCL3 [26], and IL10RB [27]. Each module's type and magnitude of influence on disease predictions is visualised in Figure 3.

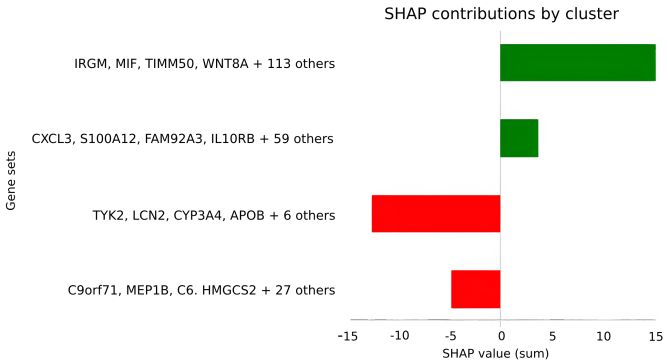

Fig. 3: Final gene modules identified as being associated with severe disease (CD deep ulcer), alongside relative contributions determined using SHAP values. Shown is a bar plot in which each gene module is characterised in terms of its influence on the model predicting the "CD deep ulcer" cluster. Each bar represents the sum of mean SHAP values associated with the "CD deep ulcer" cluster, across all genes in the module. We show the top 4 most influential genes of each module. More positive values (green) indicate greater confidence for that module predicting "CD deep ulcer". More negative values (red) reduce our confidence in a "CD deep ulcer" prediction. Consistent with the literature on IBD, we find genes like IRGM, CXCL3 and IL10RB are present in influential modules and have a significant effect on disease severity.

1) Gene Ontology (GO) enrichment analysis: We verified the biological relevance of identified gene modules through Gene Ontology enrichment analysis [30]–[32]. The CD deep ulcer 117-gene module (Supp. Material Fig. 8) showed enrichment in IBD-relevant processes including transport mechanisms, reactive oxygen species regulation, and immune responses, with statistically significant results (FDR ¡ 0.05) and high fold enrichment (¿100) for many processes. This aligns with IBD's characteristic dysregulated immune response to pathogens [33], showing inflammatory processes regulated through various cytokines and signaling pathways

[27]. Notably, this was the only module showing enriched regulation of reactive oxygen species, specifically associated with CD deep ulcers [20]. The 63-gene module (Supp. Material Fig. 9) revealed additional pathways involving TLR6 and TLR2 recognition of bacterial patterns [34], along with processes related to embryonic digestive tract development and extracellular matrix organisation [35], [36]. The 45-gene module (Supp. Material Fig. 10), associated with CD no ulcer, showed broader immune responses including fungal response and fc-gamma receptor signaling [37]. Our findings align with current literature [38]–[41] while uncovering novel pathways, particularly in embryonic development, potentially advancing our understanding of IBD mechanisms and CD subtype differentiation. Though smaller modules sometimes lacked enriched processes due to insufficient gene counts, the overall results strongly correspond with established IBD literature.

D. Class-contrastive explainability

We developed a class-contrastive method to visually demonstrate the impact of identified gene modules on disease subtypes. We had initially observed that gene expression across disease subtypes follows Gaussian distributions, with some genes showing clear up- or down-regulation in CD patients compared to controls, as illustrated by CXCL3 and MEP1B (Supp. Material Fig. 11). Exploring the effects of expression strength led to the development of our technique. To demonstrate our method, we analysed Patient 46 (CD deep ulcer; Fig. 4) by modifying their gene expression values to match control group means. When adjusting the 117-gene module, which showed the strongest positive association with CD deep ulcer (Figure 3), the patient's classification shifted from CD deep ulcer to CD no ulcer (Figure 5). Furthermore, modifying both the 117-gene and 63-gene modules (180 genes total) resulted in reclassification to the control cluster (Figure 6), suggesting these modules' significant role in severe disease manifestation. Similar effects were observed when analysing numerous other patients across the dataset. This technique provides intuitive patient-specific explanations. While serving as a proof-of-concept with potential for improvement through larger datasets, the method effectively demonstrates how specific gene modules influence disease classification and severity.

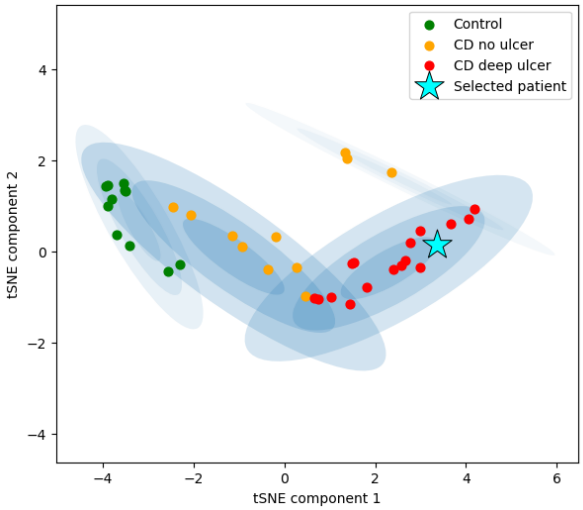

Fig. 4: Initial position of Patient 46 (CD deep ulcer) within the clustering model.

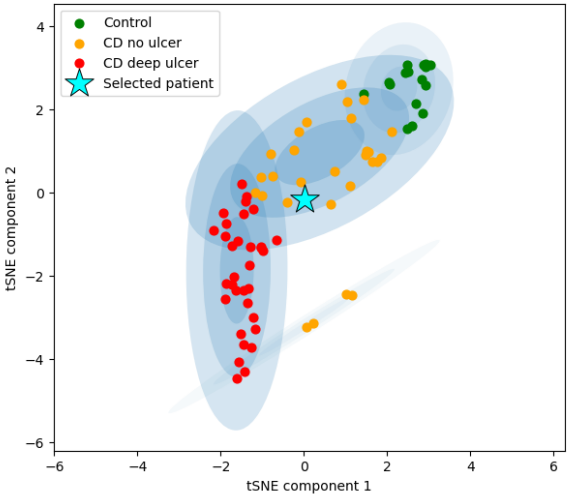

Fig. 5: 117-gene module alteration. In comparison to Figure 4, modifying the 117 genes using the class-contrastive technique results in Patient 46 (with CD deep ulcer [a severe form of the disease]) being assigned to the "CD no ulcer" cluster [a less severe form of the disease].

E. Validation with an independent cohort

To validate the generalisability of our approach, we applied our methods to an independent cohort of 210 pediatric Crohn's disease patients and 35 non-IBD controls, featuring different disease subtypes ("CD no complication" and "CD with complication"). The initial stages involved pre-processing and filtering, from which we identified 278 differentially expressed genes and applied our computational framework using a newly trained autoencoder adapted for this dataset. Results (Supp. Material Table 3; Figure 14) showed strong performance comparable to our original study. Binary classification (control vs. CD) achieved 89.2% accuracy and 93.3% F1-score using GMM with autoencoder and tSNE. Multi-class classification showed slightly lower but respectable scores (64.9% accuracy, 69.9% F1-score), reflecting the real-world challenge of distinguishing between related CD subtypes. Subsequent kernelSHAP analysis identified the most influential genes, with CXCL5, IL1B, and CXCL6 among the top three, as shown in Supp. Material Figure 15. These genes, known for their role in intestinal inflammation, showed strong influence on model predictions [42]. Several other identified genes have established links to IBD, such as FMO5's role in mucus barrier formation [43] and ALDH1A2's involvement in vitamin A signaling disruption during active IBD [44]. We also identified potentially novel targets, including WNT5A,

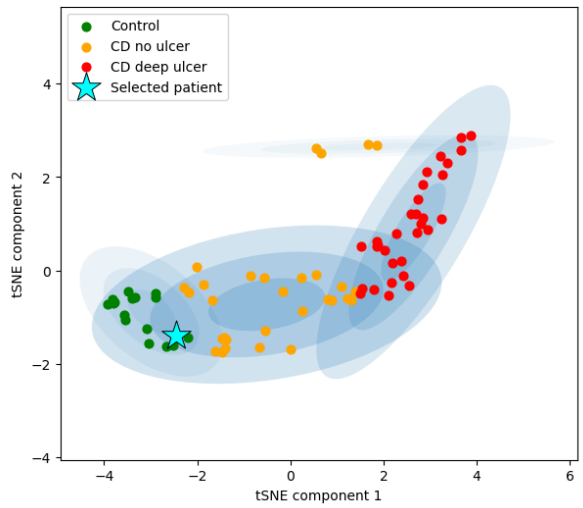

Fig. 6: 117-gene and 63-gene module alteration. Modifying both the 117-gene and 63-gene modules using the class-contrastive technique results in Patient 46 moving from the red "CD deep ulcer cluster" (Fig. 4) to the green control cluster. This suggests these modules may be involved in a severe form of CD that leads to deep ulcers.

which has recently been recognised as a potential therapeutic target. These results demonstrate our framework's ability to generalise across datasets and effectively identify both known and novel genes associated with CD subtypes.

V. Conclusions and Discussion

Compared to the state-of-the-art, one of our key contributions is in adapting kernelSHAP to account for gene correlations, enabling more accurate identification of risk genes in CD subtypes through our GMM-derived probabilistic model. This framework not only stratifies patients effectively but also captures complex relationships between expression profiles and disease subtypes.

We identify both established IBD genes (NOD2, IRGM, IL10) and potentially novel targets (including uncharacterised LOC genes). While some identified genes, such as SELE, lack strong established links to IBD, these could represent either novel findings or limitations of our approach. Through a novel data integration technique and WECR consensus clustering [2], we identify disease-relevant gene modules, with GO enrichment analysis validating their biological significance. Our class-contrastive technique also provides intuitive visual explanations of how gene modules influence disease classification at the individual patient level.

While our approach is not intended for de-novo risk gene identification, it effectively identifies disease-relevant genes and gene modules from expression data, and characterises their influence. We validate our techniques with an independent cohort in Supp. Material Section 5.5. The model-agnostic nature of our methods also makes them applicable to other domains where explainable analysis of correlated features is crucial (e.g. fraud detection, climate modelling etc.). This framework has significant potential to enhance clinical decision-making by providing interpretable insights into disease mechanisms at both population and individual levels.

References

[1] Satyam Kumar. Improve your Model Performance with Auto-Encoders, December 2021.

[2] Yongxuan Lai, Songyao He, Zhijie Lin, Fan Yang, Qifeng Zhou, and Xiaofang Zhou. An adaptive robust semi-supervised clustering framework using weighted consensus of random k-means ensemble. *IEEE Transactions on Knowledge and Data Engineering*, 33(5):1877–1890, 2021.

[3] Soumya Banerjee, Pietro Lio, Peter B. Jones, and Rudolf N. Cardinal. A class-contrastive human-interpretable machine learning approach to predict mortality in severe mental illness. *npj Schizophrenia*, 7:60, 12 2021.

[4] Wenkai Han, Yuqi Cheng, Jiayang Chen, Huawen Zhong, Zhihang Hu, Siyuan Chen, Licheng Zong, Liang Hong, Ting-Fung Chan, Irwin King, Xin Gao, and Yu Li. Self-supervised contrastive learning for integrative single cell RNA-seq data analysis. *Briefings in Bioinformatics*, 23(5), 09 2022. bbac377.

[5] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, 2017.

[6] Melvyn Yap, Rebecca L. Johnston, Helena Foley, Samual MacDonald, Olga Kondrashova, Khoa A. Tran, Katia Nones, Lambros T. Koufariotis, Cameron Bean, John V. Pearson, Maciej Trzaskowski, and Nicola Waddell. Verifying explainability of a deep learning tissue classifier trained on RNA-seq data. *Scientific Reports*, 11(1):2641, January 2021.

[7] Jin Hayakawa, Tomohisa Seki, Yoshimasa Kawazoe, and Kazuhiko Ohe. Pathway importance by graph convolutional network and shapley additive explanations in gene expression phenotype of diffuse large b-cell lymphoma. *PLOS ONE*, 17:e0269570, 6 2022.

[8] Yang Yu, Pathum Kossinna, Wenyuan Liao, and Qingrun Zhang. Explainable autoencoder-based representation learning for gene expression data. 12 2021.

[9] M. Pavageau, L. Rebaud, D. Morel, S. Christodoulidis, E. Deutsch, C. Massard, H. Vanacker, and L. Verlingue. DeepOS: pan-cancer prognosis estimation from RNA-sequencing data. preprint, Oncology, July 2021.

[10] Abdul Karim, Zheng Su, Phillip K. West, Matthew Keon, The NYGC ALS Consortium, Jannah Shamsani, Samuel Brennan, Ted Wong, Ognjen Milicevic, Guus Teunisse, Hima Nikafshan Rad, and Abdul Sattar. Molecular classification and interpretation of amyotrophic lateral sclerosis using deep convolution neural networks and shapley values. *Genes*, 12(11), 2021.

[11] Scott Lundberg. Api reference: Core explainers, 2018.

[12] Frank Emmert-Streib, Matthias Dehmer, and Benjamin Haibe-Kains. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in Cell and Developmental Biology*, 2, 2014.

[13] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature machine intelligence*, 2(1):56–67, January 2020.

[14] Kjersti Aas, Martin Jullum, and Anders Løland. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence*, 298:103502, 2021.

[15] Yan Zhang, Zhengkui Lin, Xiaofeng Lin, Xue Zhang, Qian Zhao, and Yeqing Sun. A gene module identification algorithm and its applications to identify gene modules and key genes of hepatocellular carcinoma. *Scientific Reports*, 11:5517, March 2021.

[16] Heewon Park, Koji Maruhashi, Rui Yamaguchi, Seiya Imoto, and Satoru Miyano. Global gene network exploration based on explainable artificial intelligence approach. *PLoS ONE*, 15(11):e0241508, November 2020.

[17] Xiao Ye, Yulin Wu, Jiangsheng Pi, Hong Li, Bo Liu, Yadong Wang, and Junyi Li. Deepgmd: A Graph-Neural-Network-Based Method to Detect Gene Regulator Module. *IEEE/ACM transactions on computational biology and bioinformatics*, 19(6):3366–3373, 2022.

[18] Bin Zhang and Steve Horvath. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, 4:Article17, 2005.

[19] Peter Langfelder and Steve Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1):559, December 2008.

[20] Yael Haberman, Timothy L. Tickle, Phillip J. Dexheimer, Mi Ok Kim, Dora Tang, Rebekah Karns, Robert N. Baldassano, Joshua D. Noe, Joel Rosh, James Markowitz, Melvin B. Heyman, Anne M. Griffiths, Wallace V. Crandall, David R. Mack, Susan S. Baker, Curtis Huttenhower, David J. Keljo, Jeffrey S. Hyams, Subra Kugathasan, Thomas D. Walters, Bruce Aronow, Ramnik J. Xavier, Dirk Gevers, and Lee A. Denson. Erratum: Pediatric crohn disease patients exhibit specific ileal transcriptome and microbiome signature (journal of clinical investigation (2014) 124: 8 (3617-3633) doi: 10.1172/jci75436). *Journal of Clinical Investigation*, 125, 2015.

[21] Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L. Robertson, Nadezhda Serova, Sean Davis, and Alexandra Soboleva. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Research*, 41(Database issue):D991–995, January 2013.

[22] Scott M Lundberg, Bala Nair, Monica S Vavilala, Mayumi Horibe, Michael J Eisses, Trevor Adams, David E Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, 2(10):749, 2018.

[23] Soichiro Yamamoto and Xiaojing Ma. Role of Nod2 in the development of Crohn's disease. *Microbes and infection / Institut Pasteur*, 11(12):912–918, October 2009.

[24] I. Atreya, R. Atreya, and M. F. Neurath. NF-κB in inflammatory bowel disease. *Journal of Internal Medicine*, 263(6):591–596, 2008. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2796.2008.01953.x.

[25] B. M. Fournier and C. A. Parkos. The role of neutrophils during intestinal inflammation. *Mucosal Immunology*, 5(4):354–366, July 2012. Number: 4 Publisher: Nature Publishing Group.

[26] J. Puleston, M. Cooper, S. Murch, K. Bid, S. Makh, P. Ashwood, A. H. Bingham, H. Green, P. Moss, A. Dhillon, R. Morris, S. Strobel, R. Gelinas, R. E. Pounder, and A. Platt. A distinct subset of chemokines dominates the mucosal chemokine response in inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics*, 21(2):109–120, 2005. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2036.2004.02262.x.

[27] Reza Yazdani, Bobak Moazzami, Seyedeh Panid Madani, Nasrin Behniafard, Gholamreza Azizi, Majid Aflatoonian, Hassan Abolhassani, and Asghar Aghamohammadi. Candidiasis associated with very early onset inflammatory bowel disease: First IL10RB deficient case from the National Iranian Registry and review of the literature. *Clinical Immunology*, 205:35–42, August 2019.

[28] LOC100505851 uncharacterized LOC100505851 [Homo sapiens (human)] - Gene - NCBI.

[29] Subhash Mehto, Kautilya Kumar Jena, Parej Nath, Swati Chauhan, Srinivasa Prasad Kolapalli, Saroj Kumar Das, Pradyumna Kumar Sahoo, Ashish Jain, Gregory A. Taylor, and Santosh Chauhan. The Crohn's Disease Risk Factor IRGM Limits NLRP3 Inflammasome Activation by Impeding Its Assembly and by Mediating Its Selective Autophagy. *Molecular Cell*, 73(3):429–445.e7, February 2019.

[30] Seth Carbon and Chris Mungall. Gene Ontology Data Archive, March 2023.

[31] Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, and Gavin Sherlock. Gene Ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, May 2000.

[32] Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research*, 49(D1):D325–D334, January 2021.

[33] Bani Ahluwalia, Luiza Moraes, Maria K. Magnusson, and Lena Öhman. Immunopathogenesis of inflammatory bowel disease and mechanisms of biological therapies. *Scandinavian Journal of Gastroenterology*, 53(4):379–389, April 2018. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00365521.2018.1447597.

[34] Takumi Kawasaki and Taro Kawai. Toll-Like Receptor Signaling Pathways. *Frontiers in Immunology*, 5, 2014.

[35] Alicja Derkacz, Paweł Olczyk, Krystyna Olczyk, and Katarzyna Komosinska-Vassev. The Role of Extracellular Matrix Components in Inflammatory Bowel Diseases. *Journal of Clinical Medicine*, 10(5):1122, March 2021.

[36] Cristiano Pagnini and Fabio Cominelli. Tumor Necrosis Factor's Pathway in Crohn's Disease: Potential for Intervention. *International Journal of Molecular Sciences*, 22(19):10273, September 2021.

[37] Fabian Junker, John Gordon, and Omar Qureshi. Fc Gamma Receptors and Their Role in Antigen Uptake, Presentation, and T Cell Activation. *Frontiers in Immunology*, 11, 2020.

[38] Weitao Hu, Taiyong Fang, and Xiaoqing Chen. Identification of differentially expressed genes and mirnas for ulcerative colitis using bioinformatics analysis. *Frontiers in Genetics*, 13, 2022.

[39] Mohammad Elahimanesh and Mohammad Najafi. Cross talk between bacterial and human gene networks enriched using ncRNAs in IBD disease. *Scientific Reports*, 13(1):7704, May 2023. Number: 1 Publisher: Nature Publishing Group.

[40] Chunwei Cheng, Juan Hua, Jun Tan, Wei Qian, Lei Zhang, and Xiaohua Hou. Identification of differentially expressed genes, associated functional terms pathways, and candidate diagnostic biomarkers in inflammatory bowel diseases by bioinformatics analysis. *Experimental and Therapeutic Medicine*, 18(1):278–288, July 2019.

[41] Xiaoli Pang, Hongxiao Song, Xiaolu Li, Fengchao Xu, Bingxun Lei, Fei Wang, Jing Xu, Lingli Qi, Libo Wang, and Guangyun Tan. Transcriptomic analyses of treatment-naïve pediatric ulcerative colitis patients and exploration of underlying disease pathogenesis. *Journal of Translational Medicine*, 21(1):30, January 2023.

[42] Łukasz Kopiasz, Katarzyna Dziendzikowska, and Joanna Gromadzka-Ostrowska. Colon expression of chemokines and their receptors depending on the stage of colitis and oat beta-glucan dietary intervention—crohn's disease model study. *International Journal of Molecular Sciences*, 23(3):1406, January 2022.

[43] Megan L. Schaller, Madeline L. Sykes, Joy Mecano, Sumeet Solanki, Wesley Huang, Ryan J. Rebernick, Safa Beydoun, Emily Wang, Amara Bugarin-Lapuz, Yatrik M. Shah, and Scott F. Leiser. Fmo5 plays a sex-specific role in goblet cell maturation and mucus barrier formation. April 2024.

[44] Siri Sæterstad, Ann Elisabet Østvik, Elin Synnøve Røyset, Ingunn Bakke, Arne Kristian Sandvik, and Atle van Beelen Granlund. Profound gene expression changes in the epithelial monolayer of active ulcerative colitis and crohn's disease. *PLOS ONE*, 17(3):e0265189, March 2022.