

Digenic variant interpretation with hypothesis-driven explainable AI

Federica De Paoli¹, Giovanna Nicora¹, Silvia Berardelli^{1,2}, Andrea Gazzo³, Riccardo Bellazzi^{1,2}, Paolo Magni^{1,2}, Ettore Rizzo¹, Ivan Limongelli¹, Susanna Zucca^{1,*}

¹enGenome Srl, Via Ferrata 5, 27100, Pavia, Italy

²Department of Electrical, Computer and Biomedical Engineering, University of Pavia, 27100, Pavia, Italy

³Department of Pathology, Memorial Sloan Kettering Cancer Center, 10065, New York, NY, United States

*To whom correspondence should be addressed. Email: szucca@engenome.com

Abstract

The digenic inheritance hypothesis holds the potential to enhance diagnostic yield in rare diseases. Computational approaches capable of accurately interpreting and prioritizing digenic combinations of variants based on the proband's phenotypes and family information can provide valuable assistance during the diagnostic process. We developed diVas, a hypothesis-driven machine learning approach that interprets genomic variants across different gene pairs. diVas demonstrates strong performance in both classifying and prioritizing causative digenic combinations of rare variants within the top positions across 11 cases with the complete list of variants available (73% sensitivity and a median ranking of 3). Furthermore, it achieves a sensitivity of 0.81 when applied to 645 published causative digenic combinations. Additionally, diVas leverages explainable artificial intelligence to elucidate the digenic disease mechanism for predicted positive pairs.

Introduction

Identifying disease-causing mutations is a critical task in human genetics. This is especially important for rare diseases (RDs), as ~80% of these conditions have a genetic component [1]. RDs affect >300–400 million worldwide [2]. Accurate genetic diagnosis may enhance disease management and support more personalized treatment approaches [2]. Advances in the analysis and interpretation of whole genome sequencing (WGS) and whole exome sequencing (WES) have enhanced our ability to identify genomic variants and determine the genetic causes of RDs [3–5]. Computational tools have become essential for interpreting and prioritizing these variants [6]. For those suspected of having an RD, the goal is to find the genetic cause by analyzing the genome sequence along with patient's phenotype and family history. Most RDs are monogenic, leading clinicians to search for the single mutated gene causing the disease [7]. The American College of Medical Genetics, with the Association for Molecular Pathology, has set guidelines to classify genomic variants based on factors like familial segregation and predicted variant impact. These guidelines help medical professionals identify the specific mutation responsible for a patient's disease [8].

Despite the advancement in technical capabilities and knowledge, the diagnostic yield, defined as the percentage of patients affected by a genetic disease who undergo genetic testing and have the genetic cause identified, varies between 35% and 55% depending on the disorder [1]. Overall, considering also the inaccessibility of the genetic tests, ~200 million patients live without a definitive diagnosis [5].

In contrast to the monogenic inheritance hypothesis, oligogenic models propose that the manifestation of a genetic disorder arises from the combined occurrence of mutations in multiple genes. The simplest form of this paradigm is digenic inheritance (DI), wherein the disease's origin is attributed to two mutated genes. In the DI field, distinct mechanisms have been recognized: true digenic (TD, also known as pure digenic), composite (CO, also known as modifier), or dual molecular diagnosis (DM). In the case of TD, the disease's phenotype(s) emerge when both causative genes carry mutations. Modifiers denote variants that, when present alongside causative variants, can influence disease outcomes, such as severity or age of onset. Lastly, the DM mechanism entails the co-occurrence of two distinct disorders, each caused by pathogenic mutations in different genes [9].

The DI paradigm holds significant potential for supporting the diagnosis of RDs [7]. For instance, various genes acting as digenic contributors in arrhythmogenic cardiomyopathy, a disorder typically regarded as monogenic, have been revealed [10]. Several other conditions have also been investigated in the context of DI patterns, including hypogonadotropic hypogonadism, deafness, ciliopathies, and Long QT syndrome, with causative digenic combinations also being confirmed [11]. However, estimating DI prevalence is challenging due to the complexity of gene–gene interactions and disease variability. Less than 1% of known OMIM diseases have been reported as potentially oligogenic in OLIDA, a recently released database [12], while in a large cohort study of

Received: October 28, 2024. Revised: March 3, 2025. Editorial Decision: March 8, 2025. Accepted: March 10, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of NAR Genomics and Bioinformatics.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

2076 RD-affected patients, DM were reported for 4.9% of the patients [13].

To pinpoint the causative digenic diagnosis from WGS/WES data, potentially thousands of candidate gene pairs carrying rare variants must be assessed. Recent years have seen the emergence of numerous statistical and machine learning (ML) tools that assess the impact of individual mutations [14–16], or predict variants' pathogenicity [17–19]. Some methods further incorporate the patient's phenotypic data to prioritize variants, aiming to discern the causative one among the top contenders [20]. A limited number of computational tools have been specifically designed to classify and prioritize digenic pairs [21]. VarCopp was the first published ML tool to categorize variant pairs as either pathogenic or benign, relying on 11 features grouped into two subsets: variant-level features (*in silico* predictions of the damaging effects of the variants in a pair) and gene-level features that capture both gene–gene interactions and a priori gene-level properties such as haploinsufficiency [22, 23]. Subsequently, an additional ML model named digenic effect (DE) predictor was developed to further elucidate the three digenic mechanisms [24]. ORVAL is an online platform that processes standard variant call format (VCF) files containing the proband's variants, utilizing both VarCopp and the DE predictor for digenic classification and oligogenic network analysis [25]. DiGePred is an ML approach, available as a stand-alone tool. Like ORVAL, DiGePred uses different gene–gene interaction features, such as co-expression, protein–protein interaction, and pathway similarity, and gene-level features, such as haploinsufficiency and gene essentiality. While DiGePred autonomously classifies gene pairs [26], it does not provide the variant-specific resolution that ORVAL offers. A more recent ML methodology for predicting digenic interactions at the gene level is DIEP (digenic interaction effect predictor) [27]. OligoPVP is an automatic variant interpretation tool that can be used to prioritize digenic and oligogenic instances [28]. OligoPVP is the sole tool that integrates proband's phenotypes for interpretation. OligoPVP initially prioritizes each variant based on a phenotype-driven pathogenicity score, subsequently prioritizing sets of two or more interacting genes as determined by a high-confidence gene–gene interaction network. The final score is the sum of the pathogenicity scores of all the variants involved in the combination.

ORVAL, DiGePred, and DIEP were firstly trained using known positive digenic pairs from DIDA [29], a curated dataset listing 258 pathogenic variant pairs sourced from literature.

DIDA has recently evolved into OLIDA, encompassing >1000 oligogenic variant combinations [12]. The need for expanded open-access resources concerning digenic and oligogenic variants, similar to ClinVar [30], is crucial for understanding these complex genetic interactions and for providing a valuable reference for clinicians and researchers, aiding in accurate diagnoses and treatment decisions. Additionally, it is essential to raise awareness and educate the community about digenic and oligogenic inheritance to promote collaboration and research. For this reason, there is a pressing need to establish a community platform where researchers can share candidate causative combinations and associated patients' phenotypes.

We introduce diVas, a hypothesis-driven ML tool designed for digenic variant interpretation, grounded in monogenic

ACMG/AMP standard guidelines and gene–gene interaction data. diVas can also harness the proband's phenotypic data to enhance the classification and prioritization of digenic pairs. For positively predicted pairs, diVas employs cutting-edge explainable artificial intelligence (XAI) techniques for further subclassification into the three distinct digenic mechanisms and for providing clearer explanations and interpretations of results. We have validated diVas' capability to identify the true causative pair between thousands of candidates for 11 patients affected by different genetic diseases. The sequencing data for these patients, provided as a complete variants' list in VCF standard format, were collected through different collaborations (see the “Acknowledgements” section). We also compared our tool with existing methods for digenic variant interpretation.

We applied diVas to analyze our comprehensive high-confidence, manually curated dataset of digenic combinations, and associated phenotypes. Predictions and the XAI-layer results are freely available at oliver.engenome.com. Furthermore, diVas method for digenic variant prioritization is integrated in the eVai software (evai.engenome.com), commercialized by enGenome (www.engenome.com).

Materials and methods

Dataset description

Training data collection and preprocessing

To collect data for training purposes, we curated a dataset of pathogenic digenic combinations extracted both from DIDA [29], from literature review [13, 31–33], and from an internal unpublished database. The combinations of variants extracted and curated from the DIDA database were subsequently confirmed and released in the OLIDA database [12]. Variants are reported with their genomic coordinates and associated phenotypes expressed as human phenotype ontology (HPO) terms [34], if explicitly available. For digenic pairs in the DIDA database that lacked HPO terms, we utilized phenoBERT [35] to extract phenotypes from associated publications. In positive cases where scientific publications were unavailable, we employed a prevalence-based phenotype sampling method that draws from the pool of phenotypes linked to the respective disease(s) in the HPO resource: the more frequently a phenotype appears in the manifestation of a disease, the more likely it is to be extracted when describing the digenic combination. Each causative digenic combination is associated with a median number of five HPO terms [range (1–14)]. The positive training set contains 238 unique gene pairs and 316 unique genes. At the variant level, the distribution of variant effects in the positive training set is shown in [Supplementary Fig. S1](#), with >90% having a coding effect. Taking into account both variants and HPO terms, each digenic combination in the training set is unique.

The negative dataset contains ~5000 neutral digenic combinations randomly extracted from 256 samples of 1000 Genomes Project (1KGP) data and 10 000 from 10 solved monogenic ($n = 6$) or digenic ($n = 4$) cases with the complete variants' list available (internal data). Digenic negative combinations derived from the 1KGP were paired with either a randomly sampled set of phenotypes from a DIDA-positive digenic combination or a pool of phenotypes chosen through a prevalence-based approach from randomly selected disorders. For each sequenced case, we first processed the entire

VCF file to produce all potential digenic combinations. Subsequently, if the clinical case was a solved digenic one, the causative digenic combination was removed from the list of all possible digenic combinations. Finally, both for monogenic and digenic cases, we selected a subset of the remaining combinations that included:

- (i) at least one variant with a significant pathogenic impact on the gene (7% of the subset);
- (ii) one of the variants from the causative diagnosis (3% of the subset); and
- (iii) combinations chosen at random (90% of the subset).

In real-world scenarios, the ratio of positive-to-negative digenic instances is ~ 1 in billions, a ratio that could not realistically be achieved during our training set generation. Instead, we designed an unbalanced training set, with pathogenic combinations constituting 2.5% of the total data.

Digenic combinations generation

DiVas has the capability to analyze both single nucleotide variants (SNVs) and insertion/deletion (INDEL) variants in both the GRCh37 and GRCh38 genome assemblies.

Each instance generated by diVas to represent a digenic combination comprises a minimum of two variants, which can be in a homozygous or heterozygous state across both genes. For more details, see Supplementary material (the “Digenic combinations generation” section).

Features set

The feature set employed to characterize each digenic combination is designed to encompass the effects of variants on gene functions, interactions between genes, and associations between genes and phenotypes. The detailed description of the features and the model are reported in [36]. Details on feature computation are reported in Supplementary material (see the “Model features” section and [Supplementary Table S1](#)).

Classification system

We designed a structured pipeline implementing the diVas algorithm that accepts a list of variants from the proband as its input. These variants must be provided in a VCF file, compatible with both the GRCh37 and GRCh38 assemblies. DiVas performs variant annotation, generates all potential candidate digenic pairs as previously described, and categorizes each pair as either pathogenic or nonpathogenic, relying on the probability of pathogenicity predicted by the ML model. Our classifier, tailored to a “phenotype-driven” approach, is trained using the comprehensive variant-level, gene-level, and phenotype-level feature set detailed in Supplementary material (see the “Model features” section).

The phenotype-driven model also requires as input the list of the phenotypes observed in the patient, as HPO terms. Since certain scenarios, such as prenatal testing or carrier screening, may pose challenges in accessing the patient’s phenotypic information, we also introduced a “phenotype-free” model, which relies solely on variant-level and gene-level features, while excluding phenotypic information from its input.

For the phenotype-driven model, following the prediction process, if a pair is identified as pathogenic, a Bayesian approach that harnesses XAI through (SHapley Additive exPlanations) SHAP [37, 38] is employed. This approach serves

to determine, whether the positively predicted pair adheres to a DM mechanism or not (No-DM), thereby predicting the digenic mechanism. The classification workflow for the phenotype-driven approach is shown in Fig. 1.

Digenic prediction model

We trained the ensemble ML model for digenic variant interpretation using a strategy of 500 repeated hold-out validations.

To train each of the 500 ML models included in the ensemble final model, we select 65% of data for training, 20% for validation, and the remaining 15% is held out for testing purposes. Each subset is stratified, meaning that the proportion of classes is the same (2.5% of pathogenic) as in the complete dataset. Additionally, to reduce possible circularity and over-optimistic results on the test set, at a given step, the held-out test set has no overlap in terms of gene pairs with instances belonging to training or validation set.

At each step of the 500 repeated hold-out validations, on the 65% used for the training set, five-fold inner cross-validation is performed to select the best hyperparameters (i.e. the model that results in the highest *F1* score, harmonic mean between precision and recall) for each ML classifier that we tested. Subsequently, the classifier trained with the best hyperparameters is used to predict the 20% used for validation set. On these instances, we also selected the best threshold for classification. Usually, an ML predictor classifies as positive all instances whose predicted probability is >0.5 . The 0.5 threshold can work well on balanced problems when the class distribution is the same [39]. On the contrary, we select the threshold that maximizes the *F1* score, on the validation predictions. Finally, we evaluate the performance of the models in terms of *F1* score with their optimized thresholds on the 15% held-out test set.

Both for the phenotype-driven and the phenotype-free approaches, we tested different ML classifiers, such as random forest, logistic regression, and gradient boosting.

For the diVas phenotype-free approach, the training procedure was the same. In this case, however, we decided to train the model on TD and CO pairs alone: in fact, we observed that in the case of DM, the values of the gene-level features are predictably low. This occurs because DM represents two distinct monogenic disorders co-occurring in the same patient, with no overlap in terms of involved genes or pathways. Including, DM to the training set, would have biased the model toward positive classification, even when the contribution of gene-level features is low, thus affecting the capability of the model to discriminate between causal and neutral digenic pairs.

Additionally, to investigate whether the phenotype-driven model may perform differently on distinct disorders, we performed a leave-one-phenotype-out validation: first, we selected a set of different disease categories using the ClinGen gene-based classification [40]: among all the clinical domains curated by ClinGen working groups, only categories with a minimum of 10 pathogenic combinations from our positive dataset, were included for retraining purposes (see [Supplementary Table S3](#)). Using this approach, we could evaluate 9 out of the 13 available groups from ClinGen. For each disease category, we excluded the pathogenic pairs associated with that category and the negative pairs involving the same genes from the training set. We trained the ML classifier with

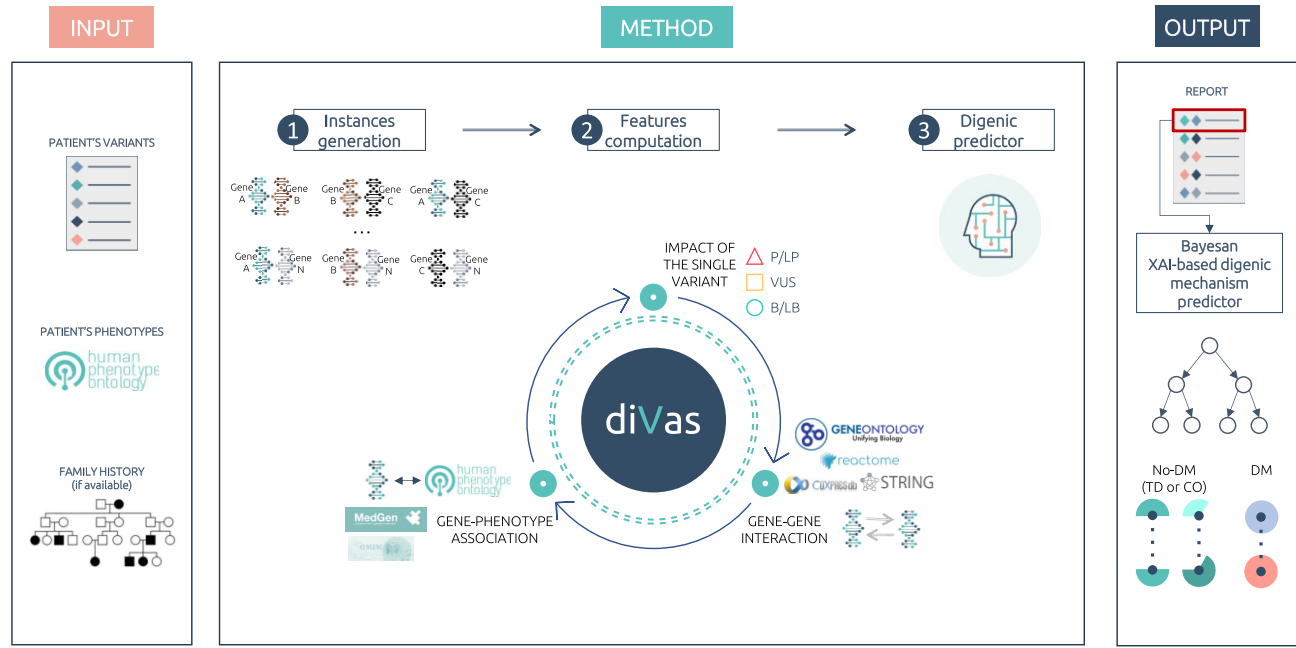


Figure 1. The phenotype-driven diVas workflow overview: diVas accepts as input the VCF files, including SNVs and INDELs, and the patient's HPO terms for proband and family members (if available). These data are then preprocessed to generate a list of potential candidate combinations. For every unique instance, a range of features is calculated. Each pair is classified as either pathogenic or nonpathogenic by the diVas model, accompanied by a probability score. These combinations are then ranked based on their predicted probabilities. Notably, for every combination identified as pathogenic, diVas conducts an analysis of the ML classifier's decision pattern, pinpointing the digenic mechanism (through the Bayesian XAI-based DE predictor module) and differentiating between DM and No-DM (both TD and CO) combinations.

the 500 repeated hold-out approach on the remaining set of variants and evaluated the performance in terms of sensitivity on the excluded pathogenic pairs. This methodology enabled the assessment of the generalization capability of diVas on instances related to disorders not represented in the training set.

XAI-based digenic mechanism prediction

After a pair is predicted to be pathogenic, understanding its specific subcategory (TD, CO, or DM) can be valuable.

In a DM patient, each mutation in different genes is independently pathogenic, resulting in two separate genetic conditions. Differentiating the digenic mechanism is crucial for guiding clinical management [41], assessing familial risk, advancing research on gene interactions, and predicting disease progression and outcomes.

One potential approach is to develop an additional ML classifier, similar to those described in previous studies [24, 41]. Along those lines, authors of previous works identified a set of informative features, such as pathogenicity of variants, gene properties, and pathways involved [24] that can differentiate between the three subcategories. However, rather than constructing an entirely new ML model based on these distinct features, we adopt a different strategy based on XAI.

XAI is an emerging topic in AI, driven by the need for more trustworthy and transparent classification systems [42]. The aim of XAI is to develop methodologies that help humans understand the classification process made by complex black-box algorithms, such as deep networks or ensemble methods. Local XAI focuses on providing explanations for individual predictions made by an ML classifier [43]. Among XAI methodologies, SHAP is a widely used framework that assigns importance scores to input features, helping to understand their contribution to the model's predictions [43].

Digenic mechanism predictor: validation approach on training set

We begin by recognizing that positive pairs falling into different subcategories should exhibit varying characteristics. For instance, in the case of DM, we expect that both variants within a pair will possess high pathogenicity scores, while the gene-gene interaction might be minimal. Conversely, in scenarios of TD or CO, we expect strong interconnections between the involved genes. Thus, we propose that the decision-making process of the classifier should diverge for DM and No-DM (TD and CO) scenarios.

We assume that the local explanation of a digenic prediction can uncover the digenic mechanism (DM or No-DM). For example, in the case of a TD pair, we anticipate that the classifier would identify it as positive due to the pronounced gene-gene interaction and phenotypic similarity. On the other hand, for a DM scenario, the decision-making process would likely place greater emphasis on the values of variant-level features. In particular, we used SHAP, a widely applied local XAI method using a game theory approach to dissect the predicted probability and assign a Shapley value to each feature. This numeric value can be interpreted as the contribution of that feature to the predicted probability [43]. To develop our XAI-based DE predictor, we compute the Shapley values for the prediction of each sample with a known DE. Since our final prediction pipeline is an ensemble of 500 classifiers, for a single instance we computed the local XAI of the classifier in the ensemble whose predicted probability for that instance is equal to the median predicted probability. Subsequently, we sum the computed Shapley values for each of the three feature categories: gene-, variant-, and phenotype-level features.

On the training set, we calculated the first percentile, the median, and the third percentile of the Shapley values for each

group. For each instance and for each feature group, we assign a categorical attribute “low,” “low-median,” “median-high,” or “high” if the sum of the Shapley values for that group is below the first percentile, between the first percentile and the median, between the median and the third percentile, or above the third percentile. Therefore, each positive predicted instance will be further characterized by three categorical features: *gene_level_contribution*, *variant_level_contribution*, and *pheno_level_contribution*, each having four different possible values (*low*, *low-median*, *median-high*, and *high*). We calculated the probabilities of being DM given each feature’s contribution to classification, and we used a naive Bayes approach to classify a new positive predicted instance as follows:

$$P(DM|X_i) \sim P(DM) * \frac{P(DM|gene_level_contribution_i)}{P(DM)} * \frac{P(DM|variant_level_contribution_i)}{P(DM)} * \frac{P(DM|pheno_level_contribution_i)}{P(DM)} \quad (1)$$

Equation 1 : Naive Bayes-predicted probability formula.

Where $P(DM)$ is calculated as the frequency of DM variants in training (28%).

To evaluate the effectiveness of the SHAP-based Naive Bayes within the repeated hold-out validation, we applied the following strategies:

1. For each seed, we calculated the SHAP values on the validation set for predicted positive samples with a DE (DM or No-DM) available.
2. On the validation set, we calculated the gene_level, variant_level, and phenotype_level contributions as the sum of the SHAP values. We then calculated the median and percentile to categorize such contributions. We calculated the probability of being DM and the conditional probabilities on such a set of data.
3. For each test sample predicted as positive and with an available DE, we calculated the posterior probability of being DM according to the Naive Bayes approach described in Equation (1). On these predictions, we evaluated different metrics, such as precision in identifying DM pairs, recall, and specificity.

Benchmarking with existing digenic interpretation tools

DiVas performances were compared with other recently developed digenic interpretation tools, namely ORVAL [25], DiGePred [26], DIEP [27], and OligoPVP [28]. While ORVAL, DiGePred, and DIEP were examined in terms of both sensitivity, specificity, and prioritization capability, OligoPVP’s evaluation focused solely on its prioritization performance due to its lack of binary classification. The primary goal of a prioritization method is to generate the shortest possible list of candidate diagnoses, ranked in order of likelihood, so that geneticists and clinicians can efficiently review the most probable ones first. It should strike a balance by achieving high recall, ensuring that true causative diagnoses are not missed, and maintaining a low false discovery rate (FDR) to minimize incorrect candidates. Ideally, the causative diagnosis should appear at the top of the candidate list. Therefore,

Table 1. Recall of the phenotype-driven and phenotype-free diVas algorithm on 500 repeated sampling iterations, stratified on different categories: TD, CO, and DM (for phenotype-driven only)

	Mean	Median	25th Percentile	75th Percentile
Phenotype-driven TD recall	0.864	0.895	0.813	0.962
Phenotype-free TD recall	0.731	0.733	0.612	0.867
Phenotype-driven CO recall	0.965	1.000	0.933	1.000
Phenotype-free CO recall	0.899	0.917	0.846	1.000
Phenotype-driven DM recall	0.910	0.917	0.857	1.000

a crucial metric to evaluate such methods is the ranking of the causative variant combination within the prioritized list. Details are provided in the Supplementary material (see the [Supplementary Table S2](#)) and the “Benchmarking with existing digenic interpretation tools” section.

Results

Digenic interpretation

Pathogenicity classification

Before analyzing the digenic dataset with digenic-specific variant interpreters, we investigated whether monogenic variant interpretation guidelines alone could support the digenic analysis. We interpreted all single variants in pathogenic pairs according to ACMG/AMP standard guidelines implemented as in our previous study [44]. Therefore, each variant is categorized in the five-tier system defined by the ACMG/AMP guidelines [8]. After interpreting each causative training variant according to the ACMG/AMP guidelines, we found that 59% of variants in DM pairs, 47% in CO, and 34% in TD are interpreted as pathogenic/likely pathogenic, and up to 16% of variants (in case of TD pairs) are benign/likely benign ([Supplementary Fig. S2](#)). These results confirm the need for the development of digenic-specific guidelines and approaches for interpretation.

Training results

On a 500 repeated hold-out validation, the final best performing model selected for digenic pathogenicity prediction, based on the evaluation of different metrics statistics (such as recall, specificity, precision, balanced accuracy, AUPRC, and F1 score), is the phenotype-driven ensemble random forest. Finally, we focus on the maximization of the AUPRC, as it is one of the most informative metrics for imbalanced problems.

DiVas demonstrates commendable performance in detecting both pathogenic (recall) and benign (specificity) combinations (Fig. 2). When performance is examined in relation to TD, CO, and DM, the average recall in predicting as pathogenic TD is slightly lower (0.864), compared to 0.965 for CO and 0.91 for DM (Table 1).

Notably, when stratified across digenic mechanisms, the phenotype-driven model exhibits superior performance compared to the phenotype-free model (see Fig. 2). For the reasons discussed above, the DM mechanism was not considered for the phenotype-free model. Specifically, the average recall for TD is 0.864 with the phenotype-driven approach,

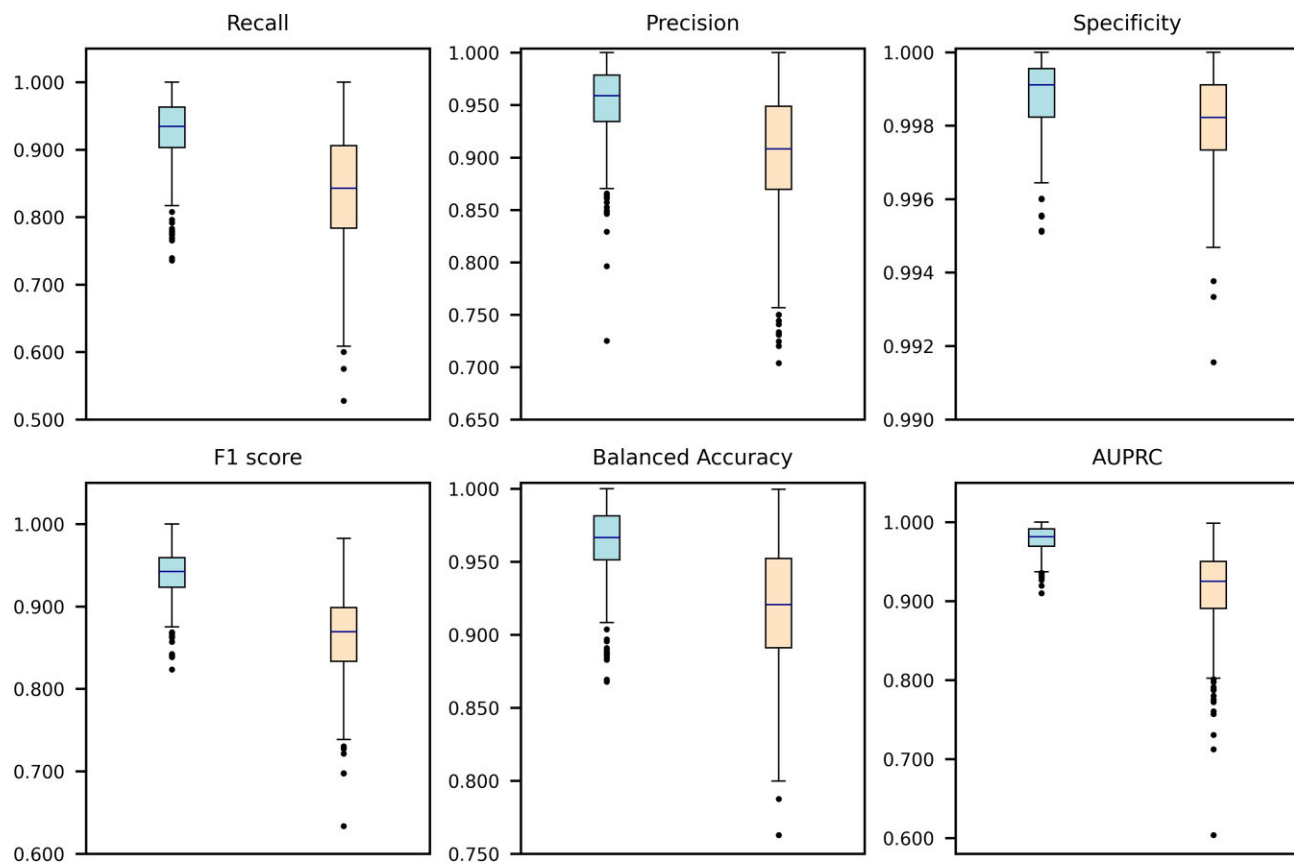


Figure 2. Boxplot of different performance metrics across 500 repeated sampling iterations: comparison between phenotype-free (on the right of each subplot) and phenotype-driven (on the left) diVas algorithms on recall, precision, specificity, F1 score, balanced accuracy, and AUPRC.

compared to 0.731 with the phenotype-free approach. On the other hand, the phenotype-driven approach achieves an average sensitivity of 0.965 for CO, which decreases to 0.899 with the phenotype-free approach. The distribution of metrics across the 500 sampling iterations can be observed in [Supplementary Figs S3 and S4](#).

The model's generalization capabilities were assessed using the “leave-one-phenotype-out” approach. Since only positive digenic pairs were stratified by the diseases they cause, performance was assessed exclusively in terms of recall. [Supplementary Table S3](#) shows the number of True Positive and False Negative pairs in nine different disorder categories. DiVas shows high recall for some disease categories, in particular, in kidney disease, inborn errors of metabolism, and ocular disorders. The lower recall (0.917) is shown on skeletal disorders. Overall, these results suggest that diVas classification is reliable even when the prediction is made on variant pairs associated with disorders not represented in the training dataset. Having a good generalization ability for populations not represented in the training data is crucial for high-stakes applications like healthcare [45].

XAI-based digenic mechanism prediction

For each predicted pathogenic combination, the Naive Bayes XAI-based digenic mechanism prediction model was applied. The conditional probabilities of being DM, given each feature-level contribution to classification, have been computed using the positive training set. For each feature subgroup (*gene-level*, *variant-level*, and *phenotype-level*), [Supplementary Table S4](#) reports the categorical value (*low*, *low-median*, *median-high*,

and *high*), representing the contribution of that feature's subgroup to classification (proportional to the SHAP values), along with the conditional probability for DM and No-DM classification, where No-DM includes both TD and CO. We observe that there is a 94% probability of a pair being classified as DM when there is a low contribution from the gene-level feature. Since gene-level features broadly represent the gene-gene interactions, this result is in line with the expectations for DM pairs, where no interaction is expected a priori between the impacted genes. On the contrary, the probability of being DM given a low variant-level contribution is low. Also in this case the XAI classification pattern aligns with expectations: variants in DM should exhibit high pathogenicity, whereas TD and CO may consist of variants that are less pathogenic when considering a monogenic hypothesis (according to ACMG evaluation) with genes that have a strong interaction ([Supplementary Fig. S2](#)). Therefore, the variant-level contribution to TD classification is expected to be low.

On the digenic mechanism prediction performed during the 500 repeated hold-out validation, results are reported in [Fig. 3](#). In median, the percentage of DM variants in the test folds used to assess the performances is 28%, in line with the general proportion of DM in the complete set.

Validation on gene panel/WES cases and benchmark analysis

DiVas has been validated on an independent dataset comprising 11 gene panel or WES samples with confirmed digenic diagnoses collected through international collaboration (see the “Acknowledgements” section). For these samples, the VCF file

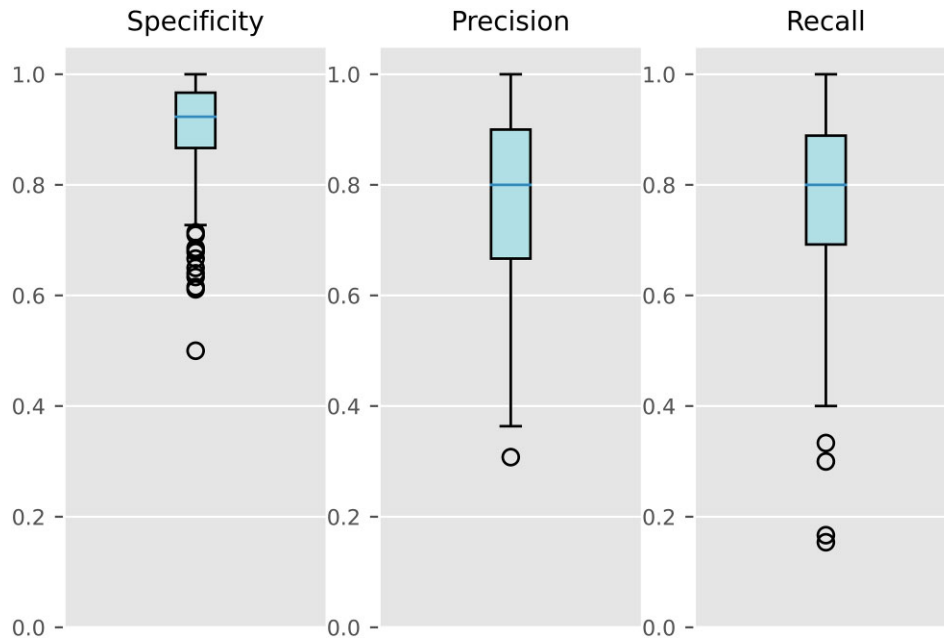


Figure 3. Boxplot of specificity, precision, and recall of the digenic mechanism predictor on 500 repeated holdout validations. Positive class is DM.

and the list of HPO terms associated with the patient were provided by the sharing group. Samples included in the validation set present heterogeneous phenotypes, ranging from skeletal to kidney disorders, hearing impairment, ocular, and cardiology abnormalities. The number of HPO terms describing patients' traits varies from a minimum of one to a maximum of four terms ([Supplementary Table S5](#)), with a median of 2. A total of six samples were single-probands, while the remaining five samples were analyzed in trio mode. For 10 out of 11 pathogenic digenic combinations, the digenic mechanism was known a priori (six No-DM and four DM).

Since all the tested tools adopt different filtering and pair generation strategies, the number of evaluated digenic instances is different. For example, ORVAL predicts all possible variant combinations, while diVas evaluates, for a given gene pair, at most four combinations with the more pathogenic variants in those genes (according to different inheritance hypotheses). On the contrary, DiGePred and DIEP do not consider variant pairs and predict the pathogenicity at the gene pair level. Finally, OligoPVP predicts the pathogenicity of each individual variant, subsequently combining variants on genes with evidence of interaction, thus potentially evaluating more than one instance per gene pair. Furthermore, diVas takes into account also family information, if available, thus reducing the number of candidate combinations to be predicted. Therefore, the number of combinations ranked by diVas, DiGePred, and DIEP is lower, potentially enhancing their prioritization ability when compared with ORVAL. For this reason, we evaluated ORVAL prioritization both in default mode and by de-duplicating gene pairs. In this case, for each gene pair, the variant combination predicted with the highest probability by ORVAL was selected. A summary about tools' performances is reported in [Table 2](#) and [Fig. 4](#), while [Supplementary Table S6](#) contains single sample's detailed statistics.

Overall, diVas shows better performance than the other tools in terms of sensitivity, median false positive rate, and median ranking of the causative digenic combination. ORVAL has a sensitivity similar to diVas. However, it predicts

nearly 30% of the evaluated instances as pathogenic, failing to adequately control the number of false positives, which is essential for a prioritization method used in realistic clinical settings. Also, DIEP shows a high number of false positives ([Supplementary Table S6](#)). De-duplicating ORVAL results by gene pair slightly improves its prioritization statistics. While DiGePred demonstrates greater specificity than ORVAL, it exhibits the lowest sensitivity, with only 2 of the 11 causative digenic combinations accurately predicted as pathogenic.

OligoPVP identifies only 2 out of 11 causative combinations, with a median of 102 prioritized pathogenic combinations identified per sample. This outcome might be attributed to the outdated resources used to assess gene-gene interactions. A deeper investigation of the causes of its high missing rate highlights that only 2 out of 11 causative gene pairs are considered as interacting by OligoPVP and thus used to build the final results. Updated gene-gene interaction resources may dramatically impact on OligoPVP's prioritization performances. For this reason, we decided to exclude OligoPVP from the benchmark analysis reported in [Table 2](#) and [Fig. 4](#).

In addition, we compared diVas and ORVAL in terms of digenic mechanism classification performance. The diVas XAI-based digenic mechanism predictor correctly classified the digenic mechanism for six out of eight digenic combinations previously identified as pathogenic by the tool, resulting in a sensitivity of 75%. ORVAL, aggregating TD and CO as No-DM to be comparable with our classification, correctly classified the mechanism for three out of six predicted pathogenic combinations ([Supplementary Table S7](#)).

Classification performance on selected cases from OLIDA database

We additionally benchmarked diVas on the manually curated high-confidence database, including 645 digenic combinations (399 unique genes for a total of 452 unique gene pairs) that were used to assess the accuracy of our tool in terms of pathogenicity and digenic mechanism prediction. More details

Table 2. Median performance of digenic prediction tools on the validation set of 11 cases

Tool	Median ranking	Sensitivity (fraction of pathogenic pairs correctly classified)	Median number of predicted pathogenic combinations (% over total number of evaluated instances)	Median number of evaluated digenic instances
diVas	3	0.73	16 (0.2%)	6656
ORVAL	154.5	0.64	3510.5 (26.15%)	13 427
ORVAL_dedup_genes	154.5	0.64	3460 (26.33%)	13 143
DiGePred	760	0.18	30 (0.5%)	5856
DIEP	346	0.55	306 (5.23%)	5855
diVas phenotype-free	5	0.73	158 (2.37%)	6656

Median ranking of the causative digenic combination and median number of predicted pathogenic and of evaluated digenic combinations are reported. Additionally, we report the sensitivity, computed as the number of true pathogenic pairs correctly classified by the tool divided by the total number of pathogenic cases (11).

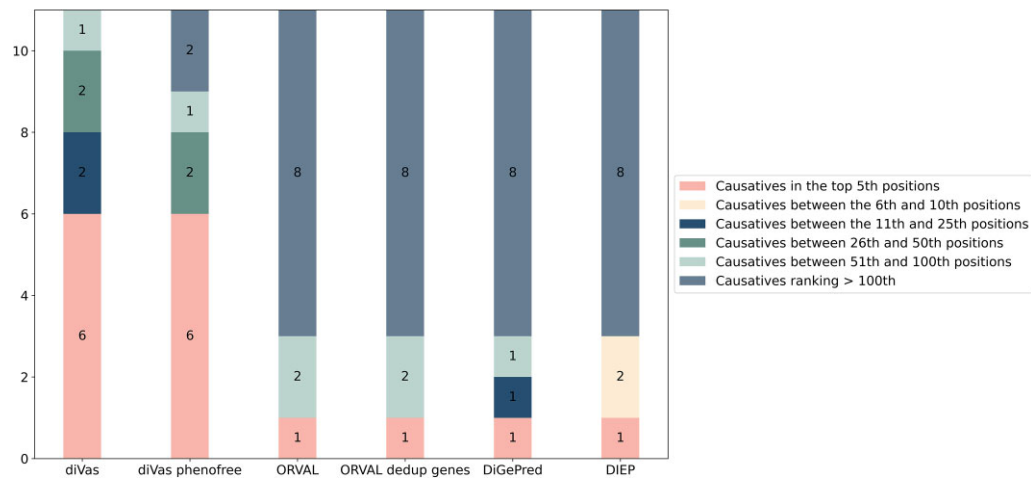


Figure 4. Performance of different tools on digenic pairs prioritization of 11 cases. For each tool, we report the number of cases for which the causative pair is ranked: (1) in the top 5th positions, (2) between the 6th and 10th, (3) between the 11th and 25th, (4) between the 26th and 50th, (5) between the 51st and 100th, or (6) above the 100th position.

are provided in Supplementary material (see the “Independent dataset of digenic combinations” section).

The other benchmarked methods were trained using DIDA/OLIDA pathogenic combinations, and the lack of control over their positive training set composition introduces potential circularity issues, which could compromise the validity of their external evaluation against OLIDA combinations.

DiVas achieves a sensitivity of 0.81 when applied to 645 causative digenic combinations using the phenotype-driven approach, while it reaches a lower sensitivity of 0.61 when employing the phenotype-free approach.

The XAI-driven digenic mechanism predictor reaches an accuracy of 76% in distinguishing TD/CO and DM, when applied to a subset of 46 causative combinations for which the digenic mechanism was known.

Oliver: sharing the dataset of pathogenic digenic combinations

A manually curated dataset of digenic pairs (not used for training purposes) is publicly available via a dedicated web app accessible at oliver.engenome.com. Due to the limited number of functionally validated digenic diagnoses available, we relied on manual revision by an expert in the field to ensure that the combinations selected for validation and released on our platform are reliable; consequently, our dataset includes both functionally validated and molecularly predicted digenic

combinations that have been published in peer-reviewed journals or confirmed through internal collaborations.

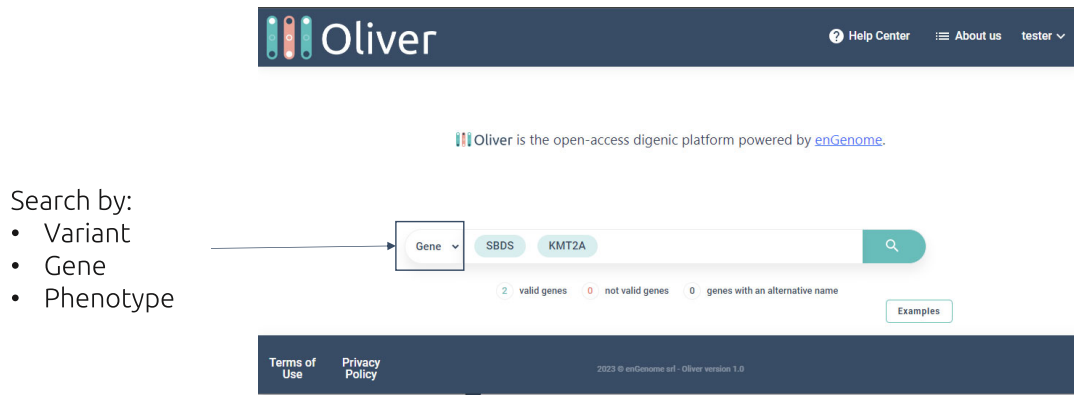
For every combination, we provide coordinates, a list of associated phenotypes, we have curated, and prediction insights from diVas and XAI-driven digenic mechanism predictor. A platform overview is shown in Fig. 5.

Discussion

Currently, the diagnostic yield of RDs is on average, far below 50% [1], and there is an urgent need to uplift this value and end the diagnostic odyssey for millions of patients that remain without a definitive diagnosis. One of the open challenges in the field of genomics is to go beyond the “one gene, one disease” paradigm that led the diagnostic genetic approach in the last 50 years [46]. Oligogenic inheritance has gained increasing attention within the ongoing challenge of identifying the causes of rare genetic disorders.

While bioinformatics and computational approaches have greatly simplified and supported the diagnostic process, the majority of available tools are still applied to each single variant in a genome [15, 17–20], without considering interactions between different elements, such as genes and their products. In the oligogenic hypothesis, the genetic cause of a disorder is attributed to the interaction of different mutated genes. As a consequence, currently available approaches for variant inter-

1



2

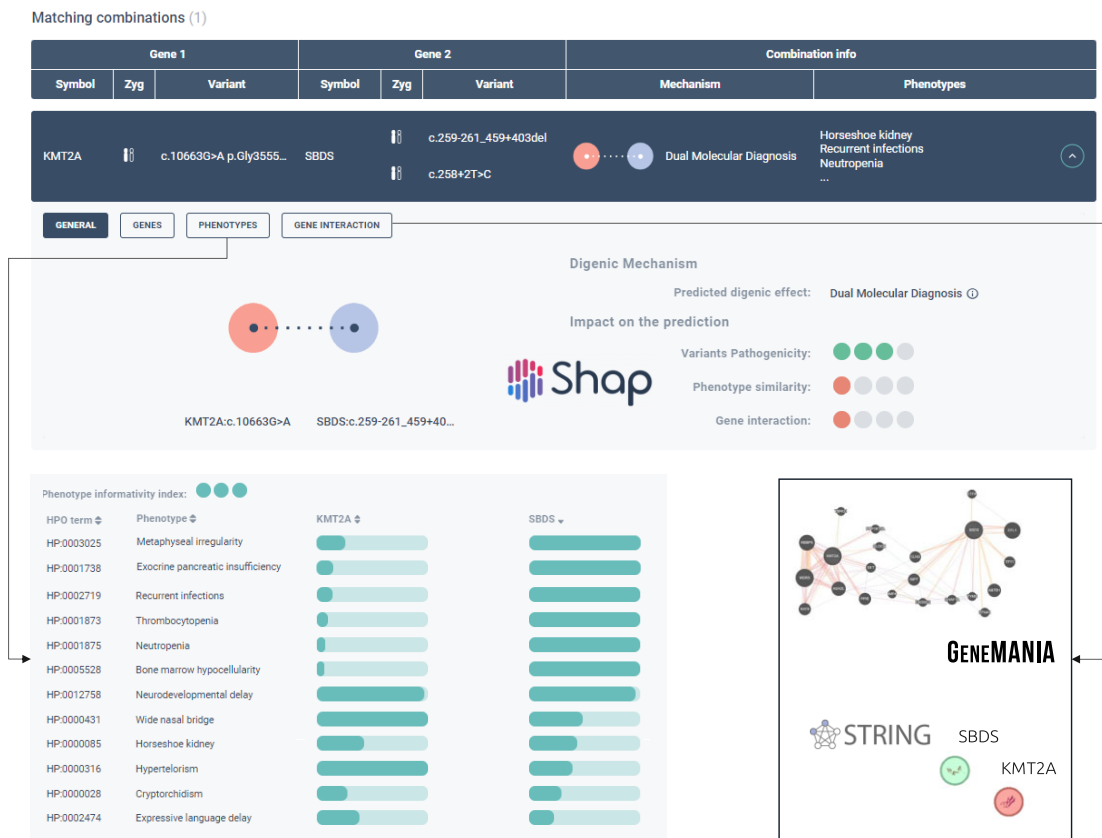


Figure 5. Oliver platform (oliver.engenome.com). Panel 1: search tab (by variant, by gene, or by phenotype). Panel 2: example of variant search and insight of a combination.

pretation are not suitable for the identification of oligogenic causative variants. The development of methodologies for the interpretation of oligogenic disorders is hampered by limited availability of bona fide pathogenic oligogenic combinations. Unlike ClinVar [30], a public repository that currently gathers thousands of interpretations under the monogenic hypothesis, the currently available database of the pathogenic oligogenic hypothesis (OLIDA) gathers around 1600 instances, the majority of them represented by digenic combinations [12]. Another fundamental aspect is that OLIDA is a manually-curated database that gathers oligogenic pathogenic instances from the literature, while ClinVar allows for the submission of clinically relevant interpretations from submitters worldwide.

Early methods focusing on digenic variant interpretation (like Orval, DiGePred, OligoPVP, and DIEP) were all trained on DIDA, the previous version of OLIDA, gathering around 250 pathogenic combinations [25, 26]. Orval recently integrated the positive training set of the ML model with the most recently released OLIDA data. Due to limited availability of bona-fide pathogenic combinations, external validation of such tools can be hampered by circularity issues [46]. The performed benchmark analysis showed that no single tool emerges as a universal standard: while Orval exhibits high sensitivity, its specificity is notably compromised. In contrast, DigePred and DIEP provide greater specificity but are constrained by their gene-level predictions, without consider-

ing the variants included in each combination. OligoPVP, currently, is the only published tool that incorporates patient phenotypes into its predictive framework; however, its sensitivity is hindered by outdated gene–gene interaction resources.

We introduce diVas, an ML tool designed for digenic variant interpretation, addressing the shortcomings of existing tools. Unique to diVas is its use of a patient's phenotypic data and family information to predict causative variant pairs; moreover, it offers explainability, aligning with the EU AI Act. The tool was trained on over 350 curated digenic instances and uses variant pathogenicity, gene interactions, and gene-phenotype associations for predictions and prioritization. In 11 cases with the whole variants' list available, diVas ranked causative pairs higher than other methods, with a median ranking of 3, improving prioritization performances compared to competitors. A more extensive benchmark validation on a larger set of digenic solved samples is required to determine, whether diVas significantly outperforms other methods, as the current sample size is insufficient for a definitive conclusion. Although, our phenotype-driven and phenotype-free models exhibit comparable sensitivity across the 11 validation cases, the phenotype-driven model achieves a lower FDR. This reduces the number of predicted pathogenic digenic combinations requiring clinical review, streamlining variant interpretation, and enhancing diagnostic efficiency. A difference in sensitivity between the phenotype-free and phenotype-driven models emerges in the validation against OLIDA combinations. This discrepancy arises because some causative variants in the OLIDA database, according to our implementation of ACMG/AMP guidelines, have a low impact on the gene (e.g. they are common in the general population). In these cases, removing phenotype-association features, combined with low scores on variant-impact features, significantly affects the final prediction, reducing the sensitivity of the phenotype-free model. The next step for diVas is oligogenic interpretation, which requires understanding gene interactions and robust datasets. With its improved performance, diVas has potential for widespread use in diagnosing patients, especially those without a molecular diagnosis.

In this context, we have released a curated dataset of digenic combinations and associated predictions performed by the diVas model on the Oliver platform (oliver.engenome.com) to advance this understanding. Oliver integrates both HPO-based standardized phenotypic information and the XAI-layer results for each pathogenic prediction.

Providing our curated data and prediction results supports researchers in their endeavors, reflecting our commitment to advancing the field. This resource is a powerful tool for hypothesis generation, findings validation, and new investigations. Beyond data sharing, we aspire to foster a community of researchers and clinicians with a shared interest in DI, promoting collaboration, discussion, and insight exchange through the Oliver platform. Our contribution is not just data presentation; it is an invitation to collaborative scientific exploration, expected to drive further advancements in understanding DI and its various implications.

Acknowledgements

We would like to extend our gratitude to the following scientists, listed in alphabetical order: Dr Alfredo Brusco and PhD student Lisa Pavinato (University of Turin, Turin, Italy); Dr Cesare Danesino, Dr Antonella Minelli, and PhD student

Ibrahim Taha (University of Pavia, Pavia, Italy); Dr Edoardo Errichiello and PhD student Mauro Lecca (University of Pavia, Pavia, Italy); Dr Paolo Gasparini, Dr Giorgia Giroto, and PhD student Anna Morgan (Institute for Maternal and Child Health–IRCCS Burlo Garofolo, Trieste, Italy); Dr Alessandro Geroldi (University of Genoa, Genoa, Italy); Dr Corrado Mammì and Dr Manuela Priolo (Hospital of Reggio Calabria Az. Ospedaliera Bianchi-Melacrino-Morelli, Reggio Calabria, Italy); Dr Nikolaos M. Marinakis, Dr Joanne Traeger-Synodinos, and PhD student Faidon-Nikolaos Tilemis (National and Kapodistrian University of Athens, Athens, Greece); Dr Marco Tartaglia (IRCCS Bambino Gesù's Hospital, Rome, Italy); and Dr Enza Maria Valente and PhD student Valentina Serpieri (IRCCS Mondino Foundation, Pavia, Italy).

We also extend our thanks to Anne Bowcock (PhD student–Icahn School of Medicine at Mount Sinai) for her invaluable support in data curation.

Author contributions: De Paoli F. (Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing-original draft, Writing-review & editing), Zucca S. (Conceptualization, Supervision, Funding acquisition, Writing-original draft, Writing-review & editing), Nicora G. (Conceptualization, Methodology, Software, Writing-original draft, Writing-review & editing), Limongelli I. (Conceptualization, Funding acquisition, Writing-review & editing), Berardelli S. and Gazzo A. (Data curation, Writing-review & editing), Rizzo E. (Supervision, Funding acquisition, Writing-review & editing), Bellazzi R. and Magni P. (Supervision, Writing-review & editing).

Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

Conflict of interest

All the authors collaborate with enGenome Srl. F.D.P., I.L., and S.Z. are full employees of enGenome Srl. R.B., P.M., E.R., I.L., and S.Z. have shares of enGenome.

Funding

The results presented in this paper were funded by the European Innovation Council Accelerator (grant number: 190164416).

Data availability

The training procedure, the final model, and the prediction script are available at <https://doi.org/10.5281/zenodo.14672280>.

The results generated using the diVas method, combined with the XAI layer, on a high-confidence, comprehensive, manually curated dataset of known digenic combinations are publicly available at oliver.engenome.com. DiVas method for digenic variant prioritization is available in eVai (eai.engenome.com), a software as a service for variant interpretation that is regulated by commercial license.

The cases with the complete list of variants used to assess the prioritization capabilities of the method are not publicly available, as they are protected by memoranda of understand-

ing signed with the institutions mentioned in the “Acknowledgements” section.

References

- 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR *et al.* 100,000 Genomes pilot on rare-disease diagnosis in health care—preliminary report. *N Engl J Med* 2021;385:1868–80. <https://doi.org/10.1056/NEJMoa2035790>
- Marwaha S, Knowles JW, Ashley EA. A guide for the diagnosis of rare and undiagnosed disease: beyond the exome. *Genome Med* 2022;14:23. <https://doi.org/10.1186/s13073-022-01026-w>
- Álvarez-Mora MI, Sánchez A, Rodríguez-Revena L *et al.* Diagnostic yield of next-generation sequencing in 87 families with neurodevelopmental disorders. *Orphanet J Rare Dis* 2022;17:60. <https://doi.org/10.1186/s13023-022-02213-z>
- Frésard L, Montgomery SB. Diagnosing rare diseases after the exome. *Cold Spring Harb Mol Case Stud* 2018;4:a003392. <https://doi.org/10.1101/mcs.a003392>
- Vinkšl M, Writzl K, Maver A *et al.* Improving diagnostics of rare genetic diseases with NGS approaches. *J Community Genet* 2021;12:247–56. <https://doi.org/10.1007/s12687-020-00500-5>
- Stenton SL, O’Leary MC, Lemire G *et al.* Critical assessment of variant prioritization methods for rare disease diagnosis within the rare genomes project. *Hum Genomics* 2024;18:44. <https://doi.org/10.1186/s40246-024-00604-w>
- Rahit KMT, Tarailo-Graovac M. Genetic modifiers and rare mendelian disease. *Genes* 2020;11:239. <https://doi.org/10.3390/genes11030239>
- Richards S, Aziz N, Bale S *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405–24. <https://doi.org/10.1038/gim.2015.30>
- Deltas C. Digenic inheritance and genetic modifiers. *Clin Genet* 2018;93:429–38. <https://doi.org/10.1111/cge.13150>
- König E, Volpato CB, Motta BM *et al.* Exploring digenic inheritance in arrhythmogenic cardiomyopathy. *BMC Med Genet* 2017;18:145. <https://doi.org/10.1186/s12881-017-0503-7>
- Schäffer AA. Digenic inheritance in medical genetics. *J Med Genet* 2013;50:641–52. <https://doi.org/10.1136/jmedgenet-2013-101713>
- Nachtegaal C, Gravel B, Dillen A *et al.* Scaling up oligogenic diseases research with OLIDA: the Oligogenic Diseases Database. *Database (Oxford)* 2022;2022:baac023. <https://doi.org/10.1093/database/baac023>
- Posey JE, Harel T, Liu P *et al.* Resolution of disease phenotypes resulting from multilocus genomic variation. *N Engl J Med* 2017;376:21–31. <https://doi.org/10.1056/NEJMoa1516767>
- Flygare S, Hernandez EJ, Phan L *et al.* The VAAST Variant Prioritizer (VVP): ultrafast, easy to use whole genome variant prioritization tool. *BMC Bioinformatics* 2018;19:57. <https://doi.org/10.1186/s12859-018-2056-y>
- Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF *et al.* Predicting splicing from primary sequence with deep learning. *Cell* 2019;176:535–48. <https://doi.org/10.1016/j.cell.2018.12.015>
- Limongelli I, Marini S, Bellazzi R. PaPI: pseudo amino acid composition to score human protein-coding variants. *BMC Bioinform* 2015;16:123. <https://doi.org/10.1186/s12859-015-0554-8>
- Nicora G, Zucca S, Limongelli I *et al.* A machine learning approach based on ACMG/AMP guidelines for genomic variant classification and prioritization. *Sci Rep* 2022;12:2517. <https://doi.org/10.1038/s41598-022-06547-3>
- Rentsch P, Witten D, Cooper GM *et al.* CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 2019;47:D886–94. <https://doi.org/10.1093/nar/gky1016>
- Ioannidis NM, Rothstein JH, Pejaver V *et al.* REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am Hum Genet* 2016;99:877–85. <https://doi.org/10.1016/j.ajhg.2016.08.016>
- Yuan X, Wang J, Dai B *et al.* Evaluation of phenotype-driven gene prioritization methods for Mendelian diseases. *Brief Bioinform* 2022;23:bbac019. <https://doi.org/10.1093/bib/bbac019>
- Okazaki A, Ott J. Machine learning approaches to explore digenic inheritance. *Trends Genet* 2022;38:1013–8. <https://doi.org/10.1016/j.tig.2022.04.009>
- Papadimitriou S, Gazzo A, Versbraegen N *et al.* Predicting disease-causing variant combinations. *Proc Natl Acad Sci USA* 2019;116:11878–87. <https://doi.org/10.1073/pnas.1815601116>
- Versbraegen N, Gravel B, Nachtegaal C *et al.* Faster and more accurate pathogenic combination predictions with VarCoPP2.0. *BMC Bioinform* 2023;24:179. <https://doi.org/10.1186/s12859-023-05291-3>
- Gazzo A, Raimondi D, Daneels D *et al.* Understanding mutational effects in digenic diseases. *Nucleic Acids Res* 2017;45:e140. <https://doi.org/10.1093/nar/gkx557>
- Renaux A, Papadimitriou S, Versbraegen N *et al.* ORVAL: a novel platform for the prediction and exploration of disease-causing oligogenic variant combinations. *Nucleic Acids Res* 2019;47:W93–8. <https://doi.org/10.1093/nar/gkz437>
- Mukherjee S, Cogan JD, Newman JH *et al.* Identifying digenic disease genes via machine learning in the Undiagnosed Diseases Network. *Am Hum Genet* 2021;108:1946–63. <https://doi.org/10.1016/j.ajhg.2021.08.010>
- Yuan Y, Zhang L, Long Q *et al.* An accurate prediction model of digenic interaction for estimating pathogenic gene pairs of human diseases. *Comput Struct Biotechnol J* 2022;20:3639–52. <https://doi.org/10.1016/j.csbj.2022.07.011>
- Boudelloua I, Kulmanov M, Schofield PN *et al.* Phenotype-driven analysis of individual genomic information to prioritize oligogenic disease variants. *Sci Rep* 2018;8:14681. <https://doi.org/10.1038/s41598-018-32876-3>
- Gazzo AM, Daneels D, Cilia E *et al.* DIDA: a curated and annotated digenic diseases database. *Nucleic Acids Res* 2016;44:D900–7. <https://doi.org/10.1093/nar/gkv1068>
- Landrum MJ, Lee JM, Riley GR *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucl Acids Res* 2014;42:D980–5. <https://doi.org/10.1093/nar/gkt1113>
- Yang R, Lu Y, Yang C *et al.* Case report: expanding the digenic variants involved in thyroid hormone synthesis—10 new cases of congenital hypothyroidism and a literature review. *Front Genet* 2021;12:694683. <https://doi.org/10.3389/fgene.2021.694683>
- Iafusco F, Maione G, Mazzaccara C *et al.* NGS analysis revealed digenic heterozygous GCK and HNF1A variants in a child with mild hyperglycemia: a case report. *Diagnostics* 2021;11:1164. <https://doi.org/10.3390/diagnostics11071164>
- Chen Q, Zheng W, Xu H *et al.* Digenic variants in the TTN and TRAPPC11 genes co-segregating with a limb-girdle muscular dystrophy in a Han Chinese family. *Front Neurosci* 2021;15:601757. <https://doi.org/10.3389/fnins.2021.601757>
- Köhler S, Gargano M, Matentzoglou N *et al.* The human phenotype ontology in 2021. *Nucleic Acids Res* 2021;49:D1207–17. <https://doi.org/10.1093/nar/gkaa1043>
- Feng Y, Qi L, Tian W. PhenoBERT: a combined deep learning method for automated recognition of human phenotype ontology. *IEEE/ACM Trans Comput Biol and Bioinform* 2023;20:1269–77. <https://doi.org/10.1109/TCBB.2022.3170301>
- Limongelli I, Zucca S, De Paoli F *et al.* Predictive method for determining the pathogenicity of combinations of digenic or oligogenic variants. EU patent. EP4309185A1 (24 January 2024), Assignee: enGenome srl.
- Caruana R, Lundberg S, Ribeiro MT *et al.* Intelligible and explainable machine learning: best practices and practical challenges. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Virtual Event, CA, USA: Association for Computing Machinery, 2020, 3511–2. <https://doi.org/10.1145/3394486.3406707>

38. Saranya A, Subhashini R. A systematic review of explainable artificial Intelligence models and applications: recent developments and future trends. *Decis Anal J* 2023;7:100230. <https://doi.org/10.1016/j.dajour.2023.100230>
39. Zou Q, Xie S, Lin Z *et al.* Finding the best classification threshold in imbalanced classification. *Big Data Research* 2016;5:2–8. <https://doi.org/10.1016/j.bdr.2015.12.001>
40. Rivera-Muñoz EA, Milko LV, Harrison SM *et al.* ClinGen variant curation expert panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. *Hum Mutat* 2018;39:1614–22. <https://doi.org/10.1002/humu.23645>
41. Kadlubowska MK, Schrauwen I. Methods to improve molecular diagnosis in genomic cold cases in pediatric neurology. *Genes* 2022;13:333. <https://doi.org/10.3390/genes13020333>
42. Guidotti R, Monreale A, Ruggieri S *et al.* A survey of methods for explaining black box models. *ACM Comput Surv* 2018;51:42.
43. Lundberg SM, Erion G, Chen H *et al.* From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2:56–67. <https://doi.org/10.1038/s42256-019-0138-9>
44. Nicora G, Limongelli I, Gambelli P *et al.* CardioVAI: an automatic implementation of ACMG-AMP variant interpretation guidelines in the diagnosis of cardiovascular diseases. *Hum Mutat* 2018;39:1835–46. <https://doi.org/10.1002/humu.23665>
45. Kelly CJ, Karthikesalingam A, Suleyman M *et al.* Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195. <https://doi.org/10.1186/s12916-019-1426-2>
46. Rimoin DL, Hirschhorn K. A history of medical genetics in pediatrics. *Pediatr Res* 2004;56:150–9. <https://doi.org/10.1203/01.PDR.0000129659.32875.84>